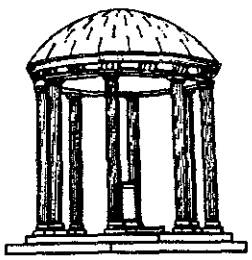
The Biological Computation of Visual Motion

TR93-036
1993

Robin Eric Fredericksen

Department of Computer Science
University of North Carolina at Chapel Hill
Chapel Hill, NC 27599-3175



The Biological Computation of Visual Motion

Robin Eric Fredericksen

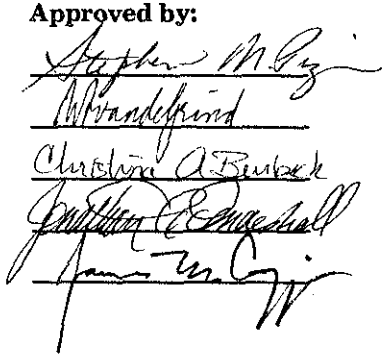
A dissertation submitted to the faculty of the University of North Carolina in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Computer Science.

Chapel Hill, 1993

Approved by:

Co-chairman

Co-chairman



Robin Eric Fredericksen. The Biological Computation of Visual Motion (Under the direction of Stephen M. Pizer and Willem A. Van de Grind.)

ABSTRACT

The mechanisms underlying the computation of visual motion in biological systems are explored. Psychophysical experiments in visual perception of apparent motion (sequences of static images as in film, video, or computer controlled displays) were performed using random dot apparent motion stimuli designed to isolate populations of biological motion detectors. The results of these experiments along with extant information in the psychophysical, neurophysiological and neuroanatomical literature are used to construct mathematical models for the distribution of biological motion detectors in the visual field, and for the mechanisms underlying the combination of motion detector outputs over visual space and time (stimulus duration). The mathematical models are shown to capture the major characteristics of the visibility of motion in humans. The detector distribution is described in terms of the visibility of spatial displacement size and image frame rate, and the dependence of that visibility on stimulus eccentricity in the visual field. The mathematical form of the distribution is a natural consequence of the mapping of visual space to visual cortex combined with a homogeneous distribution of cortical correlation distances, self-organization processes, and the availability of motion information in the visual field. The mechanisms for combination of motion detector responses over visual space and over time are modeled based on known neurophysiology. Preliminary results from simulations of the self-organization of the detector array are consistent with the psychophysical data. A comprehensive model is presented and the implications of the model with respect to the computer controlled display of motion information are discussed.

This work is dedicated to my mother,
Barbara A. Fredericksen

ACKNOWLEDGMENTS

I would like to thank:

- Steve Pizer and Wim Van de Grind and for being my Patron Saints.
- Christina Burbeck, Jonathan Marshall, and James Coggins for being my checks and balances, and for telling me when I was way off base.
- Ann Stuart for invaluable advise along the way, and most importantly, for teaching me that "the direct test is always best."
- Many people in Utrecht, especially Frans Verstraten; some for serving as psychophysical subjects, and all for providing friendship and direction to a stranger in a strange land.
- And last but by no means least, Suzanne Moon and many other people in Durham and Chapel Hill who have provided friendship, emotional support, and a sympathetic ear.

The research presented here has been supported by:

- NIH Grant PO1 CA47982,
- the UNC Cognitive Science Program,
- the International Brain Research Organization,
- and the UNC Graduate School.

TABLE OF CONTENTS

SYMBOLS AND ABBREVIATIONS

$\mathbf{D_{max}}$	largest visible two-frame spatial displacement
$\mathbf{D_{min}}$	smallest visible two-frame spatial displacement
$\mathbf{D_S}$	total stimulus duration
E	angular distance from the point of gaze (fovea)
$\mathbf{E_2}$	parameter in the cortical magnification model
$\mathbf{E_S}$	angular distance from the fovea for the center of a stimulus
H	height of a square stimulus
ISI	interstimulus interval
L	luminance of the stimulus
LSNR	luminance signal-to-noise-ratio
$\mathbf{N_D}$	the number of spatial displacements in the stimulus
$\mathbf{N_L}$	the number of coherent displacements that a dot undergoes
$\mathbf{N_{L\text{-}max}}$	the largest possible $\mathbf{N_L}$ given a specific $\mathbf{S_S}$ and screen width **W**
S^{100}_{min}	lower $\mathbf{S_S}$ side of the S100 contour
S^{100}	LSNR = 100 contour over the $\langle \mathbf{S_S}, \mathbf{T_S} \rangle$ plane
Ψ	spatial Fourier transform minus the dc component
S^{100}_{max}	upper Ss side of the S100 contour
$\mathbf{S_d}$	bi-local detector span distance
SOA	stimulus onset asynchrony
$\mathbf{S_S}$	spatial pattern displacement size of the stimulus
SSNR	spatial signal-to-noise-ratio
$\mathbf{T_d}$	bi-local detector delay
$\mathbf{T_S}$	stimulus frame duration (inverse of frame rate)
W	width of a square stimulus

Chapter 1: Introduction and background

1.1. Introduction

Light energy cannot help itself: it propagates through the universe at a constant (relative) speed. Objects in the universe that are in the light's path interact with any light energy that impinges upon them. The energy can be absorbed or re-emitted, polarized, rotated, deflected, diffracted, refracted or scattered, and each interaction creates new spatio-temporal patterns of light energy that are then available for further interactions. A consequence of this situation is that the global spatio-temporal energy pattern contains information about the objects that created it. This is important because animal visual systems have evolved to take advantage of that information source. Not surprisingly, human perception of visual events is achieved through interaction of the visual system with the incident spatio-temporal light energy patterns. Light energy, in the form of photons, is absorbed by proteins in small physical structures in our retina (called rods and cones) and transduced into electro-chemical signals. These signals are then combined and transformed in ways intended to estimate important properties of the objects around us. From an evolutionary stance the object properties are likely to be important to the survival of the individual, so the estimates should be accurate, consistent and, in some cases, rapidly performed.

This dissertation is concerned with the biological extraction of a small part of this information, namely, the motion of other objects in relation to the observer. This information is available because the spatio-temporal patterns of light energy depend on the relative positions of the light source and the object. An observer of an object is interacting with a small portion of the light energy pattern influenced by the object, and the perceived portion of spatio-temporal energy depends on the relative positions of all three: the light source, the object and the observer. If the relative positions change, that relative motion is encoded as alterations of the spatio-temporal energy pattern received by the observer's visual system. The estimation of this relative movement by the observer can be important: the motion may signal the observer that food has been found or it may signal the observed that they have been found to be food. In either case the observer probably wishes to obtain an accurate and rapid estimate of the relative movement that is taking place.

1.2. Research Motivation

There are many reasons why it is desirable for us to understand *how* these estimates of motion information take place, above and beyond the fact that they do take place. The first reason comes from the point of view of a scientist: the question of *how / why* is just plain interesting. This is a non-trivial reason because scientific research is not noted for significant monetary or lifestyle-oriented rewards. The scientist works long hours for low pay because she/he *wants to understand something*, and trivializing this reason would be disingenuous. However as a justification for society to support or allow research there must be other reasons. My explanation of why computer scientists and engineers should be interested in how our visual systems compute visual motion information is one of efficacy of engineering design.

If we, as engineers, want to build automata that can perceive motion, then what more successful examples are there than biological visual systems? Given a general visual perception task, a biological system will outperform man-made constructs of hardware and/or software every time, hands down. If the measurement domain is very narrow, for example measuring absolute luminance, or relative position in a quantitative manner, then we may be able to build superior measuring devices. However given the task of say, detecting a baseball moving against the sky, tracking it and estimating the termination point of its trajectory, our constructs fail miserably. In an evolutionary sense our construct is very vulnerable. It should therefore be a matter of course for engineers to study biological systems that successfully perform operations/calculations/measurements that are important to engineers. Some engineers may respond by stating that humans can build better devices and/or design better information processing algorithms, or by repeating the popular misconception that biological visual systems are

sloppy and imprecise in comparison to special purpose machines. My response to the first is that engineers have so far failed to prove that statement through demonstration. My response to the second is that you will not believe that if you study biological visual systems.

A perhaps more important reason is that the computer and requisite video display technology are pervading our society to an ever increasing degree. Understanding the motion measurement process performed by our visual systems can guide the design of systems in which motion plays a prominent role. Information on these displays often takes the form of motion, but if we do not understand the information transfer process, we can not present the information reliably, and accurately. A specific example we can choose is one of flight simulation technology. Without some understanding of the properties of motion perception we would not be able to build convincing simulations of flight. A more general example just coming into the mainstream is a field often called virtual reality, or virtual worlds. In this case the observer is presented with visual information designed to simulate some other environment than the one that they are actually in. The utility of the method lies in allowing the observer to move about in the world, and that motion is not convincing if the visual information is not accurate. Understanding of our motion perception mechanisms may aid us in designing the algorithms used for deciding what elements of the world to display and how they need to be displayed. This can be important because compute power is one of the primary limiting factors in establishing realism in these systems.

1.3. Research problem, approach, and contributions

The underlying thesis of this dissertation is that our ability to perceive visual motion (to extract motion information from our environment) is dictated by the properties of our visual system components and architecture. With that starting point the research presented here studies the following basic questions about our perception of visual motion.

(1) What are the distributions of motion detector properties for a given position in the visual field?

(2) How are motion detector outputs filtered over time to improve our perception of motion?

(3) How are motion detector outputs filtered over space to improve our perception of motion?

The first research step was to design an experimental measurement method. The design process was based on a model of motion perception that is described in detail in section 5.3. Experiments designed to shed light on the answers to the above three questions were then performed using the new experimental method. Where appropriate, mathematical models were derived and/or fitted to the psychophysical data that were measured. The new experimental paradigm was not always adequate for answering some questions about the detector array. As a means of proceeding further, a self-organizational simulation was designed to generate hypotheses about the relationship between the spatial and temporal characteristics of a motion detector and that detector's preferred motion velocity.

The first significant contribution to science presented in this dissertation is the novel combination of motion stimulus and threshold measurement method. Previous investigations of motion perception with apparent motion stimuli (visual motion generated by a sequence of static images) were limited in one way or another by the parameters of the stimulus that could be (or were) controlled. The new experimental method is a novel combination of previous methods that allows independent control of a wider range of stimulus parameters than has previously been possible. Most importantly, the method allows control over the spatio-temporal correlation content of the stimulus and the stimulus duration.

Significant contributions to our understanding of motion perception result from the use of the new experimental paradigm. The combination of the psychophysical and self-organizational simulation results allows unification and refinement of previous research, as well as resolution of apparently conflicting conclusions that resulted from the use of differing experimental methods. The experiments presented here provide (1) a

more detailed measurement of the perceivable range of spatial displacement sizes as a function of frame rate and eccentricity of the stimulus in the visual field, (2) a more precise measure of the mechanism underlying temporal integration (combination of motion information over stimulus duration), and (3) a verification and extension of previous research concerning spatial summation (the combination of motion information over visual space).

The unification of previous results is embodied in the formulation and/or refinement of models for (1) and (2) above. The models are based on known neuroanatomy and neurophysiology, and capture the behavior of the psychophysical data in a way that is consistent with the structure of the biological system. Finally, the analytical nature of the integration of these two models provides a basis for future testing and verification.

1.4. Dissertation Outline

Chapter 2 explores the visibility of random dot apparent motion stimuli and the dependence of that visibility on spatial displacement size of the pattern, frame rate, and eccentricity of the stimulus in the visual field. The chapter concludes by presenting a model for the variation of the maximum and minimum visible spatial displacement sizes based on known visual anatomy.

Chapter 3 discusses the dependence of the visibility of the same stimulus on spatial displacement size, frame rate, and displacement count (the number of frames - 1) as a means of exploring the improvement of visibility of the motion over time. The conclusion reached is that total stimulus duration is a better indicator of temporal integration of motion information than is displacement count. A first order model of temporal integration based on total stimulus duration and temporal contrast sensitivity is proposed.

Chapter 4 first discusses the temporal integration of motion information as a function of total stimulus duration. The chapter then goes on to discuss the role of temporal contrast sensitivity and response threshold in the temporal integration mechanism. A second order model of leaky integration of detector output is proposed for temporal integration based on the response properties of directionally selective cortical neurons, and the results of the simulation are compared to the psychophysical threshold data.

Chapter 5 discusses the relationship between temporal integration and spatial summation (the combination of individual detector outputs across visual space). The conclusion is that the two mechanisms are independent under all the tested conditions, and that spatial summation follows the theoretical signal processing limit of probabilistic summation under *some* of the tested conditions. There is also evidence in the psychophysical data, confirming previous conclusions, that the receptive fields of the biological motion detectors are elongated in the direction of motion.

Chapter 6 proposes a model for the distribution of visible spatial displacement sizes and frame rates across the visual field using the work presented in chapters 3-5. The model is constructed based on (1) the mapping of visual space to visual cortex implemented by the projection of retinal ganglia (via the lateral geniculate nucleus) to visual cortex, (2) general architectural properties of the retina and visual cortex, and (3) the self-organizational capabilities of biological nervous systems.

Chapter 7 concludes by presenting an integrated model describing the visibility of the stimulus used in the experiments presented here as a function of spatial displacement size, frame rate, stimulus duration, position of the stimulus in the visual field, and height and width of the stimulus in the visual field (a square display). The validity of that model is discussed and a range of verification and extension experiments are proposed.

Each of chapters 2-5 are written in the form of a self-standing, multiply authored journal paper. Chapter 2 has been accepted in substantially similar form to *Vision Research*. Chapter 3 has been submitted in substantially similar form to *Vision Research*. Chapters 4 and 5 will both also be submitted to the same journal, again, in substantially similar form. References to the work contained in each of these chapters are in a

dual form, including both the chapter number *and* the name under which the paper was accepted or submitted, or under which it will be submitted.

1.5. Background

The information presented here is intended for two distinct audiences, each with somewhat different expected backgrounds: computer scientists/engineers, and visual scientists. The two groups have different arsenals of tools and different bags of cultural language (jargon). Because the experimental work is visual psychophysics, the primary mode of the presentation is biased toward the visual scientist's domain. The central chapters (chapters 2-6) are designed so that they can be read independently, and without delay for the visual scientist familiar with current motion perception research. However they do build on each other, so sequential reading will provide the most continuity for the reasoning process. Sequential reading will also reinforce the ideas because of the redundancy required to make the chapters independently readable. Each central chapter includes its own bibliography and appendix, but a comprehensive bibliography is also included.

A computer scientist/engineer will generally be most interested in what the experimental results mean to the engineering process, and the conclusions chapter is aimed at interpreting the experimental results and the resulting mathematical model in that light. However some part of that audience will want to understand why the experimental data support or imply a particular conclusion or part of the overall model that I present in the final chapters. Sections 5.1 and 5.2 of this chapter are intended to provide a cursory introduction for those computer scientists/engineers who want to venture into the land of visual science. Finally, both audiences will be interested in the sections 5.3-5.6 of this chapter because they provide an explanation of the experimental approach/philosophy that guided the design of the experimental work.

1.5.1 The perception process

For a cogent discussion of the study of perception and the surrounding philosophy I recommend the introductory chapter of Perception, by Sekuler and Blake. Unfortunately I only have the space here to provide a sketch of the concepts required for understanding the work presented in the following chapters. The sketch is intended to provide the basis for understanding the main tool of a branch of experimental psychology called psychophysics.

What is the perception process? How do we define it or think about it? If I ask a person whether she is cold or warm, hungry or thirsty, upside down or right side up, whether the wind is blowing, or whether she can hear a particular sound, she can normally provide you with a response. By asking these questions I am casting that person in the role of a signal detector and asking that they make some decision about the detection or measurement of a particular signal. An abstract form of this process is shown in figure 5.1.

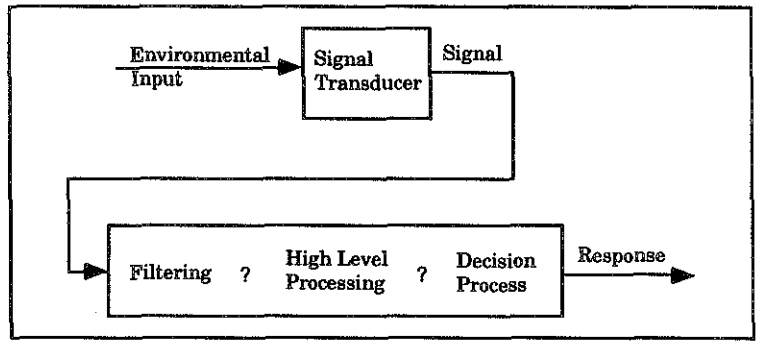


Figure 5.1
From a biological point of view the perceptual process can be described by a block diagram of a sequence of steps that begin with energy in the environment. That energy is received by the animal's sensory system and converted into a neural signal that is compatible with the biological signal processing hardware. The signal is filtered to various degrees in order to extract important information about the environment. The filtered signals, or measurements, are then used by other biological mechanisms to make decisions for the organism.

In a human being the environmental input corresponds to any energy in the environment that can be measured by a human sensory system. The

signal transducer box represents a different set of physiological components for each sensory system. The array of signal transducers that human beings are blessed with is formidable, and is the box that we know the most about. The filtering mechanisms can be considered as the neural machinery that is most peripheral in the system. That is, the processing stages that are between the actual signal transducers and the higher areas of the brain represent the filter mechanisms. In the visual system the filtering mechanisms include all of the early visual system from the horizontal cells in the retina, to the retinal ganglion cells, to the lateral geniculate nucleus, up to and including the first stages of the visual cortex. We understand various aspects of the next stages of the filtering process, but we know little or nothing about the mechanisms between the filter box and the decision-making box.

The biological approach to studying perception consists of invasion of the biological system with probes and measurement devices to test the instantaneous states of individual physiological components. The psychophysical approach to studying perception consists of manipulation of the input so that we can test the data transformation, called a *transfer function* in electrical engineering, represented by the boxes between the environmental input and the response. The biological approach is almost exclusively performed in non-human animals of various kinds because of the danger to the subject: the invasion process can often result in permanent changes to the subject's physiology. The psychophysical approach is taken less in animals than in humans because of the difficulty in controlling the form of the subject's response: it is much easier to train a human being to make a decision, although some limited kinds of experiments can be made using trained animals. The general result of this dichotomy is that we know more about human psychophysics than we do about animal psychophysics, and we know more about some animal neurophysiology than we do about human neurophysiology. Making connections between these areas of knowledge is difficult, but it is the path to enlightenment for understanding biological computation, and it is the path that I have tried to walk. In psychophysical experiments we control the input to the system and then ask the decision maker (the subject) to make a decision. We then study the input/output relationship and try to estimate the transfer function. That transfer function is implemented by components that are hidden inside the boxes and that are connected together in a specific architecture. The research work presented here is primarily visual psychophysics accompanied by mathematical or computational models designed to capture that transfer function, with the form of the model based on what we know about the neurophysiology (the functioning of the components hidden inside the boxes) and the neuroanatomy (the general architecture, or layout of the components inside the box).

1.5.2 Perception quantification

We can present a human being with all kinds of input and then ask that person to respond in some fashion, and we can build a large database of input/output response pairs. However without some framework for quantifying the decision making process and for analyzing the input/output relationship we cannot gain much insight into the nature of the transfer function. In this section I try to explain the basic structure of the framework used to understand what we measure when we perform psychophysical experiments.

The important system parameters that we deal with are the *signal*, added *noise* and the transfer function. The signal is the attribute (energy source) of the environment that we want to measure, and the noise is any interfering process, or source of energy. For example, most everyone has driven down a road in their car and lost reception of a radio signal because they traveled too far from the transmitter. The signal strength is converted to acoustic energy that we listen to. However the farther from the transmitter that we travel, the more hiss, crackle and pop we get in addition to the original transmission. This latter is the noise that interferes with the signal in which we are interested[1].

Under real world conditions there is not a great deal of uncertainty (most of the time) in our perception of events in the environment. However, an interesting situation to consider is when the environment contains both

[1]Be aware that we are using the terms signal and noise in an abstract sense. Some noise sources could be considered signal if we ask a different question of the subject, but we treat any interference as something that might be filtered out in order to improve the reliability of the signal detection/measurement process.

signal and noise sources because our performance under noisy conditions can tell us something about the perception process. The transformation process represented in figure 5.1 may be composed of (or represented as) discrete stages, each of which can include the introduction of an additional source of noise (but not usually additional signal). In a linear system[2] we can represent this as a signal + noise input function, an output function, and a single transfer function block in between because noise added internal to the process can be transformed to the input and considered as a part of the input noise. The signal and noise can be thought of as statistical processes that in general have parameters (mean, standard deviation, etc.) that result in overlapping output values of the noise, and signal + noise distributions. Figure 5.2 shows a one dimensional random variable (y) example. The value of the variable y can represent the acoustic energy emerging from your car radio speakers. The probability density distributions of the noise ($p_n(y)$), and signal + noise ($p_{s+n}(y)$) distributions are shown as overlapping.

Because the two distributions overlap, there can be no perfect separation of the signal from noise. There are many values of the variable y that have a likelihood of being produced by either noise alone, or the signal with the added noise. There is corresponding uncertainty in our signal measure or signal detection reliability. The detection task can be cast as equivalent to picking a value of y, labeled as $y_{decision}$ in the figure, and deciding that the signal is detected if it is above that value and not detected otherwise. If we choose a value for $y_{decision}$ and test the detection process a number of times with the signal present and the signal absent, we will get a percentage of detection responses determined by the distributions of both signal and noise. If we increase the strength of the signal (e.g. increase the mean of the signal distribution as shown in figure 5.2) with respect to the noise source, then we are decreasing the uncertainty of the signal detection or measurement process by reducing the overlap of the two distributions. More generally, there are situations in which the signal is present and detected (true positive), present and not detected (false negative), not present but detected (false positive), and not present and not detected (true negative). These are equivalent to the areas under the distribution curves to either side of $y_{decision}$.

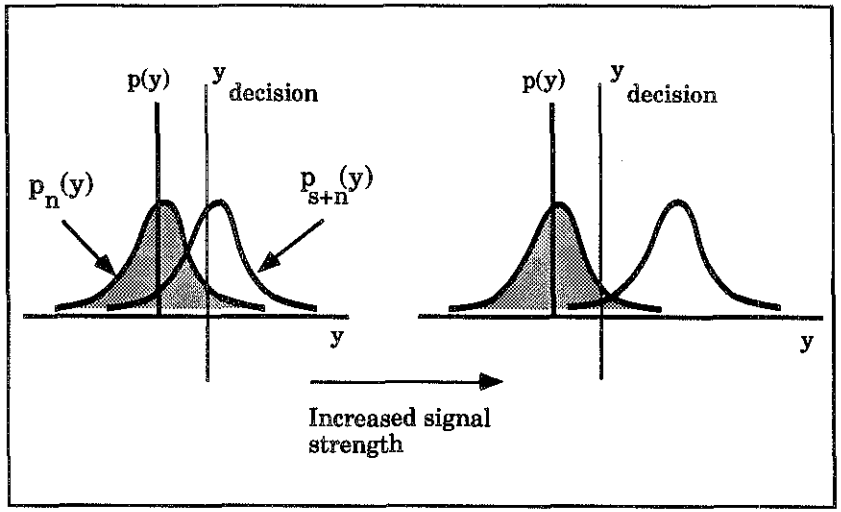


Figure 5.2
The signal and noise can be thought of as statistical processes that in general have parameters (mean, standard deviation, etc.) that result in overlapping output values of the noise, and signal + noise distributions. The figure shows a one dimensional random variable (y) that can represent the magnitude of acoustic energy emerging from your car radio speakers (a function of the radio signal and noise). The probability density distributions of the noise ($p_n(y)$), and signal + noise ($p_{n+s}(y)$) distributions are shown as overlapping. Because the two distributions overlap there can be no perfect separation of the signal from noise. The signal detection task can be cast as equivalent to picking a value of y, labeled as $y_{decision}$ in the figure, and deciding that the signal is detected if it is above that value, and not detected otherwise.

The fraction of responses in each category in a situation where the signal is presented 50% of the time can be calculated from the distributions in figure 5.2 as

$$\text{TruePositive} = \frac{1}{2}\int_{y_{decision}}^{\infty} p_{s+n}(y)dy \qquad (5.1)$$

[2] Although the biological system is not generally a linear one, we make the assumption that it can be approximated as linear so that we can employ some useful tools that emerge from linear systems theory.

(5.2) $$\text{FalseNegative} = \frac{1}{2}\int_{-\infty}^{y_{decision}} p_{s+n}(y)dy$$

(5.3) $$\text{FalsePositive} = \frac{1}{2}\int_{y_{decision}}^{\infty} p_n(y)dy$$

(5.4) $$\text{TrueNegative} = \frac{1}{2}\int_{-\infty}^{y_{decision}} p_n(y)dy$$

When we are testing the ability of a human being to detect a visual signal, the signal is contained in the visual stimulus and the observer is usually asked to detect that signal. If we perform such a measure for a human being, the resulting relationship between the signal strength and the probability of detection is called a psychometric function. (Note that the term signal refers to the information to be detected while the term stimulus generally refers to the sum of all inputs.) An idealized version of such a function is illustrated in figure 5.3 as the curve labeled Exp. 1.

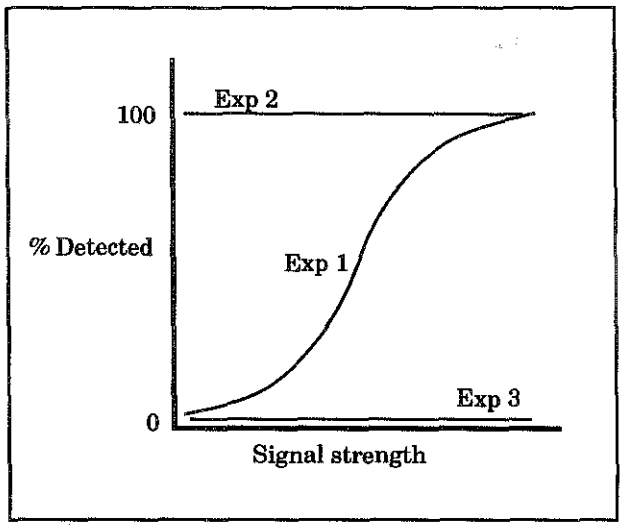


Figure 5.3
The relationship between signal strength and probability of detection that results from measuring the signal detection performance of a human being is called a psychometric curve. Exp 1 represents an unbiased signal detection process while Exp 2 and Exp 3 represent differently biased signal detection processes.

The distribution of true positive, false negative, false positive and true negative responses underlying the psychometric function can depend on a wide range of conditions, including practice, fatigue, and motivation. Each of these factors influences the position of the decision criterion represented by $y_{decision}$ in figure 5.2 For example, suppose we set up an experiment in which a subject is asked to detect a visual spatial pattern against an obscuring background pattern. If we run the first experiment with the instruction to simply detect the stimulus, we will measure a psychometric function with some distribution of true and false positive and negative responses (curve Exp 1). If we then run a second experiment, but tell the subject that she will receive $1 for each correct detection but no penalty for false detections, then the subject could decide to respond positively every time and thus maximize the monetary return (curve Exp 2). The subject moves $y_{decision}$ to a value of y=0. The subject will always detect the is signal when it is presented, but will also always indicate a false detection when the signal is absent. There will be no false negatives, and no true negatives. If we change the instructions for a third experiment by adding a penalty of strong electric shock for false positives, this again changes the decision criteria of the subject. She may decide that no money is worth the possibility of shock and respond negatively all the time, again changing the psychometric function by moving $y_{decision}$ to a value of y = infinity (curve Exp 3). For threshold measurement purposes we want to maximize true positives and true negatives, and at the same time minimize the false positives and false negatives. If each of these objectives is equally important, one solution is to set $y_{decision}$ to a value defined as $p_{s+n}(y) = p_n(y)$ as shown in figure 5.4.

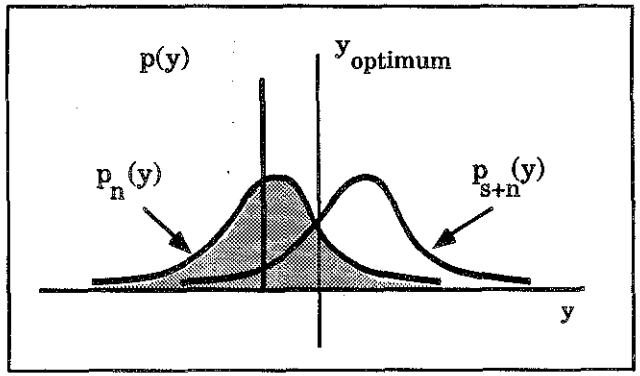


Figure 5.4
For threshold measurement purposes we want to maximize true positives and true negatives, and at the same time minimize the false positives and false negatives. If each of these objectives is equally important, one solution is to set $y_{decision}$ to a value defined as $p_{s+n}(y) = p_n(y)$.

Psychophysical measurements are therefore generally done by measuring the percentage of correct detections (true positives) using a subject motivated to detect the stimulus correctly (but without, of course, any punishments for incorrect detections!) If we measure a psychometric function under such conditions we obtain a curve (usually) something like the curve labeled Exp. 1 in figure 5.3. An idealization of such a psychometric function is called an absolute threshold curve, shown as curve b in figure 5.5.

The absolute threshold represents the absolute detectability of the signal in the stimulus for the various values of the selected stimulus parameter. An alternative method for threshold measurement uses two parameters of the stimulus rather than one. Thresholds can be measured relative to the value of an obscuring stimulus variable, for example the amount of noise in the radio signal, by selecting a percent-correct-signal-detection level. A threshold is then measured using a statistical process: the stimulus parameter of interest is held constant while manipulating the threshold variable to obtain the desired percent-correct detection level. This method provides an iso-percent-correct curve in terms of the obscuring variable rather than the threshold curve in figure 5.5. It is this latter type of threshold that used in the experiments presented here.

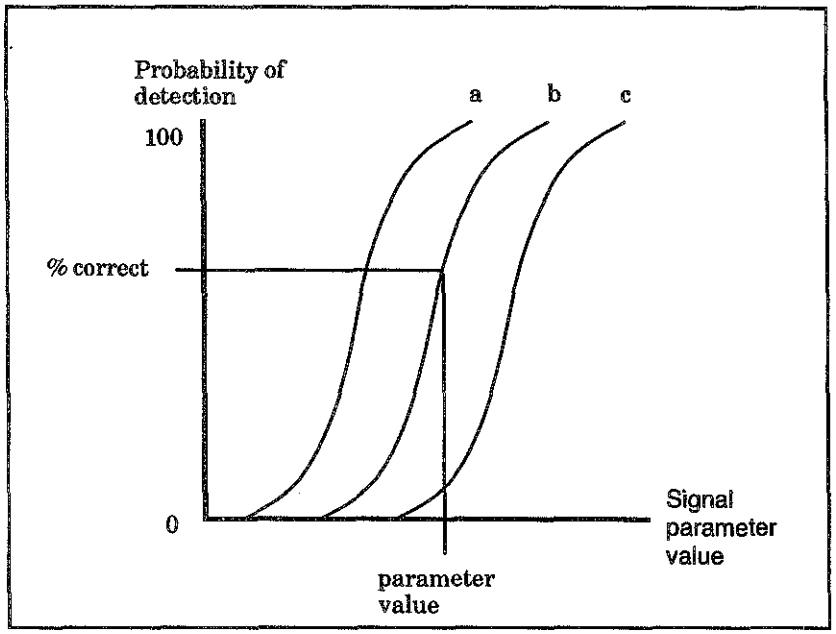


Figure 5.5
A filter is generally intended to separate the signal from the combined signal + noise input, thereby improving the reliability of the detection/measurement process. The result is equivalent to an increase of the signal strength in relation to the noise strength. Any improvement produced by the filtering process is represented by a shift of curve b to the left (curve a). Alternatively, if the magnitude of the noise increases then the curve is shifted to the right (curve c).

A filter is generally intended to separate the signal from the combined signal + noise input, thereby improving the reliability of the detection or measurement process. The result is equivalent to an increase of the signal strength in relation to the noise strength. Any improvement produced by the filtering process is represented by a shift of curve b to the left (curve a). Alternatively, if the magnitude of the noise increases then the curve is shifted to the right (curve c). Shifts of the curve result in changes of the parameter threshold value for a given percent correct threshold level.

The work presented here is concerned with how the threshold value is affected by different parameters of the stimulus. To investigate this

psychophysical threshold levels were measured for a range of stimulus parameter values (the independent variables). It is the form of the dependence of the threshold level on the independent variables that provides insight into the signal detection process and the filtering mechanisms used in that process.

1.5.3 A populational model of motion perception

The statistical nature of the threshold level measurement process generally means that a reliable estimate of a psychophysical threshold is tedious for the subject and time consuming for all those involved. It is therefore advantageous to have a specific set of tests in mind when selecting the set of independent variables to test and the range of values to test for each independent variable. Beginning with a model of the detection process can provide such guidance during experimental design. In this section I describe the model that was used in the design of the experiments presented here.

The goal of this research is to gain some understanding of how human beings extract motion information from the environment. Neurophysiological evidence tells us that the psychophysical percept of motion is based on the integration of the output of a large number of motion detectors, each of which detects motion in some portion of the visual field (called the detector's *receptive field*). The entire visual field is covered by these detectors, or alternatively, every position in the visual field is served by some set of detectors. The psychophysical percept is thus a measure of the populational response, or the entire set of detectors that respond to the stimulus. That response includes the detectors that reliably detect the motion signal as well as those that are responding at the same time but *not* to the signal. The latter group can be thought of as representing obscuring noise, and their responses can result from spontaneous firing of neurons (intrinsic noise, added internal to the detection equipment), or response to anomalous or conflicting motion information in the visual field (noise added to the stimulus before presentation to the detection machinery). The populational nature of the psychophysical percept means that any psychophysical measurements are a combination of (1) properties that emerge from the populational characteristics of the individual detectors, as well as (2) properties of individual detectors that are strong enough or widespread enough to still be visible after the integration of detector responses. Any model of the psychophysical perception process is thus necessarily a populational one. The populational model used herein is shown in figure 5.6.

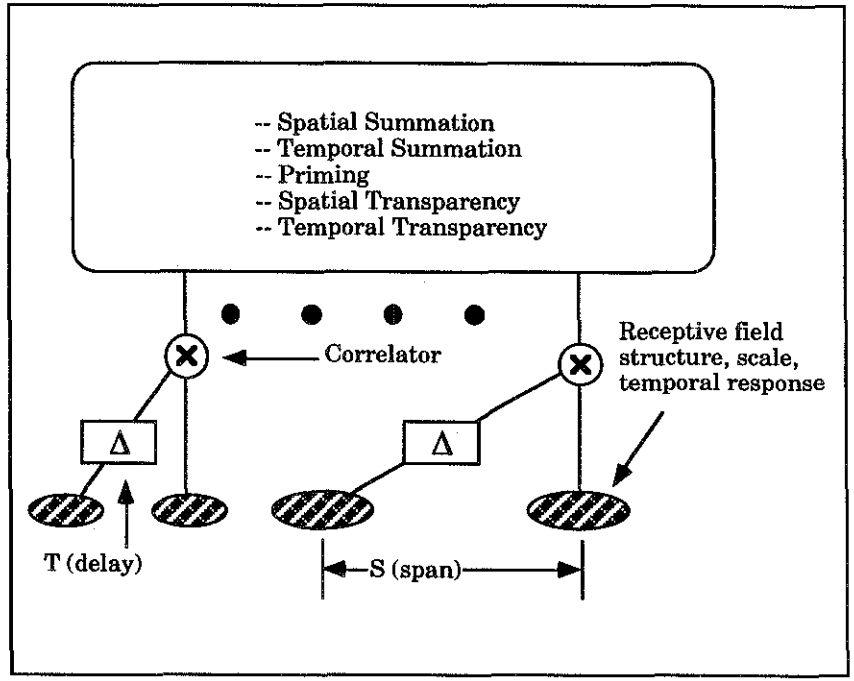


Figure 5.6
The populational model of motion detection is constructed from a front-end array of bi-local detectors that have preferred spatial displacement sizes determined by the distance between receptive field centers and a preferred visual velocity, or image frame rate determined by the differential delay of the receptive field outputs before correlation. Two bi-local detectors are shown here that differ in their receptive field size and span. The output of the array of detectors is used in higher level computations.

The model is based on the hypothesized presence of an array of spatio-temporal correlation devices which we refer to as bi-local detectors (following terminology proposed by A.J. Van Doorn [W. A. Van de Grind, personal communication]). The bi-local detector is based on the ideas of Reichardt (1961). The model used here is an adaptation of a model

developed by Reichardt as an explanation of motion perception in fly vision (his model contained counter-balanced versions of the individual detectors shown in our figure).

The basic instantiation of the model that we employ here is shown in figure 5.6. The components of the model are two spatial receptive fields, a delay component and a multiplier component. The operation of the model is very simple: the two receptive fields are structured to detect a particular class of spatial pattern in the visual field. The output signals of these components are differentially delayed and input to the multiplier (correlation) component. The output of the correlation device will signal motion in the visual field if an acceptable pattern moves across the two receptive fields in the correct direction at a speed defined by the spacing between the receptive fields and the differential delay. It is obvious that this very basic device is not perfect. For example, it will respond to flashes of the pattern at the preferred image frame rate. Elaborations and extensions of the basic idea have been proposed to remove such ambiguous or incorrect responses (Adelson and Bergen, 1985; van Santen and Sperling, 1985). However all the proposed models have a preferred spatial displacement size and a preferred temporal delay. In figure 5.6 these are represented by the distance between the receptive field centers and the delay component. We therefore show the most basic version of the concept as proposed by Reichardt.

Any given part of the visual field is served by some population of these detectors with various individual values of differential delay, receptive field center separation, receptive field size and structure, and receptive field temporal response. A visual stimulus presented in that portion of the visual field will therefore evoke a populational response from the detector population, as discussed above. The model in figure 5.5 assumes that these responses are fed to higher level computations, some of which improve the detectability of the motion in the stimulus through combination of motion detector outputs over visual space and stimulus duration. The effect of these combination mechanisms is a leftward shift of the absolute threshold curve for motion perception represented by curve b in figure 5.5.

1.5.4 Apparent motion displays

Exploration of human motion perception using psychophysical thresholds requires the use of a probe stimulus, and computers and their associated display devices provide a convenient method for producing such stimuli. We can create computer-generated image sequences that mimic the motion of real objects under a wide variety of conditions without having to manufacture a wide range of physical objects or create actual changes in the environmental conditions themselves. This utility is balanced somewhat by the additional complications introduced by the discrete nature of the motion information. The images are a sequence of snapshots of a "real" object in motion that are displayed for a period of time, the frame exposure duration, and then replaced by the next image. The motion of the object is actually a series of instantaneous spatial translations separated in time. Our visual systems nonetheless interpret the information as continuous, smooth motion if the stimulus parameters are within well known limits. This kind of motion is therefore called *apparent motion*. Note that this effect has been used for quite some time in film, video and television, and computer graphics displays. These media are all based on a sequence of static images that can provide motion information.

Although all apparent motion stimuli are based on the display of a sequence of static images, there are some minor differences between the methods used to present them. Films are stroboscopically presented sequences of static images, but television and computer graphic images are presented by drawing the image on the screen from right to left and from top to bottom. Scan-line-based presentation differs from computer graphics presentation in that the scan-line-based displays have horizontal scan lines that are analog while computer graphic displays can have discrete pixels throughout (e.g., picture tube masks) *or* analog horizontal scan lines with less discrete pixel boundaries (e.g., Trinitron tubes). Moreover, old video recording technology produced the recorded image by scanning the visual field of view in the same manner as it was displayed (newer technology uses charge-coupled-device imagers that sample each point in the visual field simultaneously, just like in photography). The scanning approach, whether as part of the recording process or as part of the display process, produces small spatio-temporal

distortions in the apparent motion information (with respect to the spatio-temporal energy that our eyes would have gathered had they been present to view the original scene). The distortions are relatively small for a single scan line, or for a local region of the patterns that are presented . A single scan line on a 90 Hz monitor using 256 scan lines is drawn in less than 40 microseconds, so any biases produced for the perception of horizontal motion should not affect psychophysical results.

Our visual systems evolved in the presence of visual motion information representing smooth and continuous displacement of objects through the world. If we consider the spatio-temporally discrete nature of apparent motion stimuli carefully, we can use them to discover properties of human motion perception.

1.5.5 Manipulable parameters of apparent motion stimuli

There are many parameters of an apparent motion stimulus that must be considered and controlled. The most general parameters, in that they apply to any kind of visual stimulus, are the position of the stimulus in the visual field (**E** or $\mathbf{E_c}$), the shape and extent of the stimulus (height (**H**) and width (**W**)), the average luminance of the stimulus (**L**), and the ambient lighting conditions. In addition, the bi-local detector model as shown in figure 5.6 indicates that successful motion detection is dependent on the spatial structure contained in the stimulus: the stimulus must activate both receptive fields of the detector. A random dot pattern contains a wide range of spatial frequency components and should therefore stimulate most or all motion detector receptive fields within the region of visual space covered by the visual stimulus (Blakemore and Campbell, 1969). Moreover, it is important to de-couple object recognition (form perception) from the more basic extraction of motion information. Random dot patterns also serve to de-couple form perception from motion perception. The experiments presented here therefore employ moving random dot pattern stimuli.

The spatial structure (Ψ) in the image can be designated as the Fourier transform of the pattern excluding the DC component (which is the mean luminance, **L**). This notion of spatial structure includes in it the average contrast of the stimulus (the relative magnitudes of the spatial frequency components) as well as the variation of that contrast across the stimulus (visual space). The stimulus contrast is also an important parameter because it influences the magnitude of the receptive field outputs. Note that the spatial structure of the objects in the image can also be changing over time. In a "real" object this could correspond to deformations or rotations of the object in motion, and/or changes in relative position of the observer and the object.

There are other parameters that are particularly important in our apparent motion stimulus. The first is the spatial displacement size between image frames ($\mathbf{S_S}$). This represents the distance traveled between camera snapshots by the object being simulated, or by the random dot pattern if we wish to exclude the influence of object recognition. The size of this displacement is an important parameter in our investigation and can be thought of as corresponding to the distance between receptive field centers ($\mathbf{S_d}$) of the motion detectors in the model.

The next parameter is the frame exposure duration ($\mathbf{T_S}$, or the inverse of the image frame rate), or total time that each image is available as input to the visual system. Manipulation of this parameter allows us to investigate the temporal properties of the portion of the motion detector array responding to the stimulus. The temporal response of a detector will depend on the temporal response properties of the individual receptive fields and the differential delay ($\mathbf{T_d}$) of receptive field outputs.

Another important stimulus parameter is total stimulus duration ($\mathbf{D_S}$). In the experiments presented here this is defined as the number of spatial displacements ($\mathbf{N_D}$, or image frame count plus one) in the apparent motion display times the frame exposure duration.

A final stimulus parameter that can be manipulated to explore the temporal properties of the motion detecting system is the interstimulus interval (ISI). The image frames containing the spatial structure representing objects in motion can be interleaved with image frames without spatial structure, and perhaps with different mean luminance

values. All experiments presented here did not explore the influence of ISI because it could not be manipulated in the experimental setup.

1.5.6 Measurement goals

The final element required for understanding and interpreting the research presented here is a model of how the apparent motion stimulus parameters interact with the populational model of motion detection and the higher level processes to determine a psychophysical threshold. The populational motion detector response is a function of a number of parameters, including spatial structure in the stimulus (Fourier transform minus the DC component; call it Ψ), mean luminance (**L**), average stimulus contrast (**C**), image frame duration ($\mathbf{T_s}$), stimulus duration ($\mathbf{D_s}$), stimulus position and size in the visual field, etc. It is possible for each of these variables to change over time but we leave them as implicit functions of time.

$$R_{signal} = f(\Psi, T_s, D_s, L, C, etc...) \tag{5.5}$$

The populational response must be large enough to be reliably detected above the noise processes masking it (see figure 5.2). We can think of this requirement as the ratio of signal to noise power being larger than some constant which will be considered as the signal-to-noise ratio (SNR) threshold.

$$\frac{R_{signal}}{R_{noise}} \geq SNR_{threshold} \tag{5.6}$$

We can also write this as minimum R_{signal} value required for signal detection.

$$R_{signal} \geq R_{threshold} \tag{5.7}$$

$R_{threshold}$ is determined, as discussed above, by the desired percent-correct level in the detection process. An R_{signal} value exactly equal to $R_{threshold}$ meets the threshold requirement. Either of these equations is a useful relation defining the required response with respect to any of our signal parameters, assuming that $R_{threshold}$ does not change as the signal changes. If we hold all but one signal parameter constant, for example the signal contrast, then we have

$$g(\Psi, T_s, D_s, L, etc...) \cdot h(C) = R_{threshold} \tag{5.8}$$

or a required signal contrast value defined as

$$h(C) = \frac{R_{threshold}}{g(\Psi, T_s, D_s, L, etc...)} \tag{5.9}$$

The contrast response function (h(**C**)) is separable from f(...) (a very desirable situation) if the response to the contrast does not depend on the other signal parameters, i.e., does not change as the other parameter values are changed. This response function determines the shape of the measured threshold curves for any selected independent variable, and is determined by the mechanisms implementing the transfer function. The mechanisms themselves are implemented by individual components (neurons) in a specific architecture (neuroanatomy). Thus the form of the psychophysical threshold curves is a result of the components and architecture of the physical system and provides insight into the transfer function implemented by the system.

In the experiments presented here the psychophysical threshold values were measured through manipulation of individual contrast values of the signal and noise inputs. The independent variables that were investigated include angular distance of the stimulus from the point of gaze, height and width of the stimulus (a square display area), spatial displacement size, image exposure duration, stimulus duration, image frame count and average stimulus contrast. The experimental method was a two-alternative-forced-choice procedure in which the subject was presented a stimulus that contained both noise and coherent motion. The observer's task was to decide the direction of the coherent motion.

It is important to explain the relationship of this task to the simple detection task discussed in section 5.2. It is assumed that, as in spatial-luminance contrast-sensitivity experiments, directional *discrimination*

is possible at motion *detection* thresholds. That is, as soon as the signal can be detected, the observer can also make some discrimination about the stimulus. The experimental task in the experiments presented here was based on *directional motion detection* in order to reduce the influence of a subject's decision criterion. However, the task also requires *discrimination* of motion directions at some level: the observer must decide (for stimuli near threshold) whether the stimulus looked more like either leftwards or rightwards motion. This is a result of using a motion noise mask to determine a threshold: motion noise is motion in all directions, including leftwards and rightwards. At the level of the motion detectors, the leftward and rightward motions are *detected* separately. The observer then "compares" (see discussion in section 5.1, and figure 5.1) the output of the two motion detector populations (left-detectors and right-detectors) and makes a discrimination decision.

Bibliography

Blakemore, C., and Campbell, F.W. (1969). On the existence of neurones in the human visual sytem selectively sensitive to the orientation and size of retinal images, *Journal of Physiology*, **203**, pp. 237-260.

Chapter 2*: The visibility of apparent motion

R. E. Fredericksen, F. A. J. Verstraten and W. A. van de Grind

Abstract

A bi-local detector array model was assumed to describe the functional performance of monocular motion perception. Distributions of model parameters were measured in human vision at several positions in the visual field. The stimulus paradigm was designed to measure directional motion perception thresholds for individual combinations of spatial displacement and temporal delay in random dot apparent motion stimuli. The resulting data support previous results on perceivable spatial displacement limits in human vision but also indicate that both minimum and maximum perceivable spatial displacement thresholds in human observers have a similar dependence on temporal delay. This dependence changes with eccentricity in the visual field in a qualitatively similar manner but by quantitatively different factors. A description of possible biological properties of the bi-local detector population is presented that may explain how detection of spatio-temporal pattern displacements can be performed by a single system. Such a model also predicts that minimum and maximum perceivable spatial displacement thresholds should scale with visual field eccentricity in a manner consistent with our results.

2.1 Introduction

The initial stage of visual motion processing has been modeled using the bi-local detector paradigm since Reichardt (1961). The mathematical properties of this model class have been explored quite extensively with respect to system decomposition (Poggio and Reichardt, 1973), spatio-temporal energy filter characteristics (Adelson and Bergen, 1985) and equivalence classes of models (van Santen and Sperling, 1985). The bi-local detector model provides a useful functional description (skeletal model) of visual motion perception performance and can be used to analyze the results of and to motivate the design of psychophysical and neurophysiological experiments. Attending to a model during the stimulus design phase is required if the results of the experiment are to be analyzable within a definite and explicit framework. In such a theory-driven approach the stimulus must be designed to measure, control or manipulate all relevant parameters of the model under investigation.

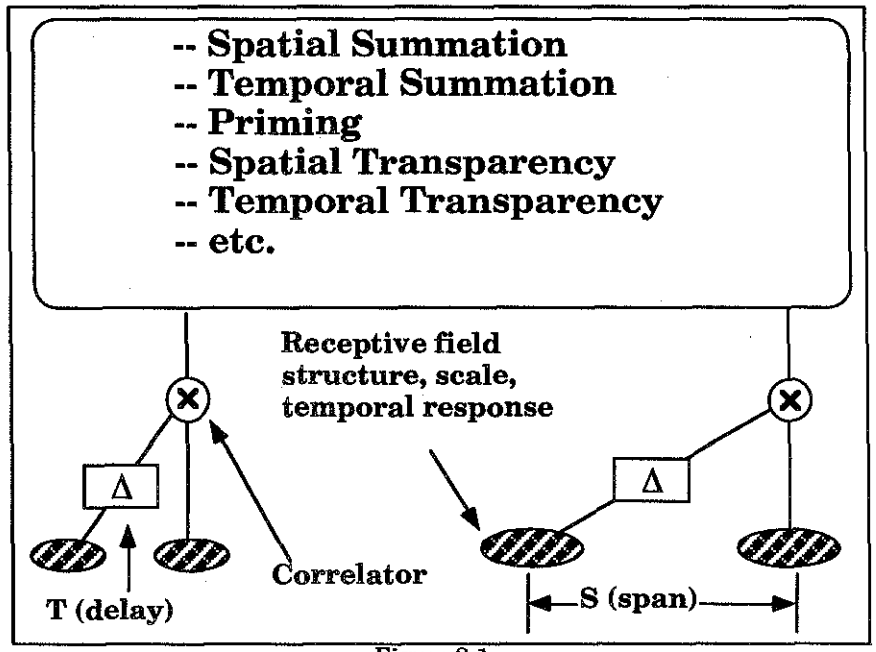


Figure 2.1
The functional performance of the human visual motion perception system is modeled as an array of bi-local detectors whose outputs are fed to later computational stages. The detector array can be characterized by distributions of detector span (distance between receptive field centers) and delay (differential delay of the receptive fields) at each position in the visual field. The receptive fields send their output to the correlation component of the detector, and can be described by their spatial structure, scale and temporal response.

* The material in this chapter will appear in the journal *Vision Research* in almost identical form under the title "Spatio-temporal characteristics of human motion perception" by the above authors.

The class of model in figure 2.1 represents an outline of what we believe to be a reasonable set of preconceptions about the functional operation of monocular motion perception in a human visual system. We assume that the operation of the motion detection system can be described by a front-end array of correlational devices, which we have represented as bi-local detectors, whose outputs are used to compute higher level (more global) information. The bi-local detectors are assumed to be characterized by the parameters depicted in the figure: (1) receptive field structure (sensitivity to spatial structure in the visual image), (2) distance between receptive field centers on the retina (span, or $\mathbf{S_d}$), and (3) differential temporal delay of the output of the receptive fields (delay, or $\mathbf{T_d}$). (The parameters **S** and **T** are used throughout the paper. The subscript **d** represents detector characteristics and the subscript **s** represents stimulus characteristics.) The output of the detector is generated through correlation of the temporal activity of the receptive fields and is input for the higher level computations. We have employed this model in the design of a stimulus used to explore the distribution of span and delay parameters of the detector array at a given position in the visual field, and the manner in which these distributions change with stimulus eccentricity in the visual field (**E**). The results reflect and encompass several findings of other authors but also show how the spatial displacement ($\mathbf{S_s}$) thresholds depend on the temporal delay ($\mathbf{T_s}$, or the interval between spatial displacements) of the stimulus. This dependence changes with **E** in a qualitatively similar manner for both maximum and minimum perceivable $\mathbf{S_s}$, but by quantitatively different factors that are explicable via inspection of the anatomy of the visual system. Interpretation of the results of these experiments in terms of the chosen model is facilitated by the measurement paradigm.

2.2 Stimulus Selection

The design of the motion stimulus is very important. This section explains the reasoning behind the selection of various stimulus parameters. The final stimulus configuration was selected so as to resolve conflicting results from two motion perception experiment stimulus types: two-frame random-dot-pattern motion and multiple-frame random-dot-pattern motion. We begin with the model of the motion perception process shown in figure 2.1 and then derive the stimulus parameter values required to resolve the previous experimental conflicts.

The important parameters of the front end bi-local detector array of the model in figure 2.1 are receptive field scale (extent on the retina), structure (center/surround, n^{th} derivative of a Gaussian, Gabor, etc.), and temporal response, as well as detector span (distance between receptive field centers), and delay (differential delay incurred outside of the receptive field temporal response). The form of this model leaves the location of other postulated/measured properties of the motion perception system to take place in a later, more global stage which integrates and evaluates the outputs of the ensemble of bi-local detectors. These higher level processes may include spatial and temporal summation (van Doorn & Koenderink, 1982a, 1982b, 1984; Nakayama & Silverman, 1984; Snowden & Braddick, 1989a, 1989b, 1991), tracking, or priming (McKee & Welch, 1985), as well as other cooperative/competitive interactions (Ball & Sekuler, 1979,1982; Williams & Sekuler, 1984; Williams, Philips and Sekuler, 1986; Brown & McKee, 1989; Bischof and Groner, 1985; Bischof and Di Lollo, 1990). There may be some disagreement about the location of such higher level processes in the visual system, but we choose to place them at a level higher than the bi-local detector (spatio-temporal filter) stage of motion processing.

A random dot pattern with a high dot density provides a stimulus containing a wide band of spatial frequencies and allows for activation of motion detectors tuned to a given span size across a wide range of receptive field sizes and structures. Such a stimulus has also been shown to exclude positional judgments (Nakayama and Tyler, 1981). Manipulation of the $\mathbf{S_s}$ and $\mathbf{T_s}$ content of the motion stimulus allows investigation of the available ranges of $\mathbf{S_d}$ and $\mathbf{T_d}$ at different values of **E**. The calculation of motion information takes place at the correlation stage of the model. The correlation stage is the location in the model that is used in our threshold measurement (discussed below). The last stage groups together a number of higher functions. The temporal properties of the receptive fields and the temporal summation of the outputs of the detectors will be intermixed, but the latter can be controlled through the number of displacements in the random dot stimulus. To explain this we first define our terminology. The *stimulus duration* will be referred to in terms of the number of spatial

displacements in the stimulus (N_D). The terms *frame* and *frame-rate* will be used to refer to a single spatial pattern presentation, and to the value of $1/T_S$, respectively. Thus N_D is calculated as the number of frames minus one. Coherent displacement means displacement of a group of dots by a constant (equal for all dots) spatial vector between frames, while incoherent displacement means displacement of each dot by an independent, randomly determined spatial vector between frames. The number of coherent spatial displacements that an individual dot undergoes during a stimulus presentation will be referred to as *dot life time* (N_L). We consider here only deterministic dot life times (all dots have the same, fixed, life time).

The original single spatial pattern displacement ($N_D = 1$) experiments for measurement of maximum detectable spatial displacement (D_{max}) (Braddick, 1973) were hindered by a less than optimal motion stimulus (Nakayama & Silverman, 1984, Snowden & Braddick, 1989a). Later experiments designed to account for these effects employed random dot patterns with N_D in the range of 2-8 and indicated that a maximum detector response requires N_D in the range of 4-6 for some $<S_S,T_S>$ combinations (Snowden & Braddick, 1989b). Increasing N_D increases the maximum perceivable S_S of a random dot pattern (Nakayama & Silverman, 1984; Snowden & Braddick, 1989a), presumably by allowing detectors of large S_S values (detectors with large S_d) to reach a maximal response. Using $N_D >> 1$ should allow the response of the target detector population to reach a maximum value and thereby obtain a competitive advantage during interactions with detectors stimulated by spatial noise.

Spatial summation can also affect the threshold for a given detector population. Scaling of the screen size in a direction *orthogonal* to the direction of motion affects thresholds by changing the number of detectors available. Scaling of the screen size in the *same* direction as that of the motion can change the range of S_d values (and thus the classes of detectors) that are activated by the stimulus. The difficulties in selecting the proper method for stimulus scaling when measuring thresholds at non-zero E has been well summarized [Drasdo, 1991]. Because it is not our primary goal to establish a *quantitative* model for the threshold gradient it is expedient to follow the convention of scaling the stimulus size in the periphery according to the value $M_S = (1 + E/E_2)$, where E is stimulus eccentricity in the visual field, and E_2 is the eccentricity in the visual field at which M_S is doubled. M_S is the dimensionless linear stimulus magnification factor, the inverse of (and thus compensating for) the (normalized) cortical magnification factor. With this definition of M_S the cortical magnification factor $M = M_f/M_S$, where M_f is the cortical magnification factor at the fovea in mm/deg.

Researchers have previously investigated both S_d and T_d characteristics of the bi-local detectors using single ($N_D = 1$) and multiple ($N_D > 1$) displacement count apparent motion stimuli with both limited ($N_L < N_{L\text{-}max}$) and unlimited ($N_L = N_{L\text{-}max}$) dot life time paradigms, where $N_{L\text{-}max} = \min[N_D, W/S_S]$ if W is the stimulus width. It has been suggested that the results of motion perception experiments with $N_L > 1$ are confounded through the simultaneous stimulation of multiple detector populations. For $N_L = 2$ in a stimulus using a spatio-temporal displacement value of $<S',T'>$, for example, there is strong motion information at spatio-temporal displacement values of $<S',T'>$ and $<2S',2T'>$.

This becomes clear if the population of bi-local detectors in figure 2.1 are mapped onto the <S,T> plane as shown in figure 2.2a. Any point on the plane represents an equivalence class of detectors tuned to $<S_d,T_d>$, while a line through the origin represents all detectors tuned to a single "average" speed $S_d/T_d = a$. (The term "average" loses meaning for large T_S values.) A motion stimulus with $N_L > 1$ provides motion information at equally spaced positions $\{<k*S_S,k*T_S>$, with k an integer such that $1<=k<=N_L\}$ in the <S,T> plane due to spatio-temporal correlation between frames that are not adjacent in time.

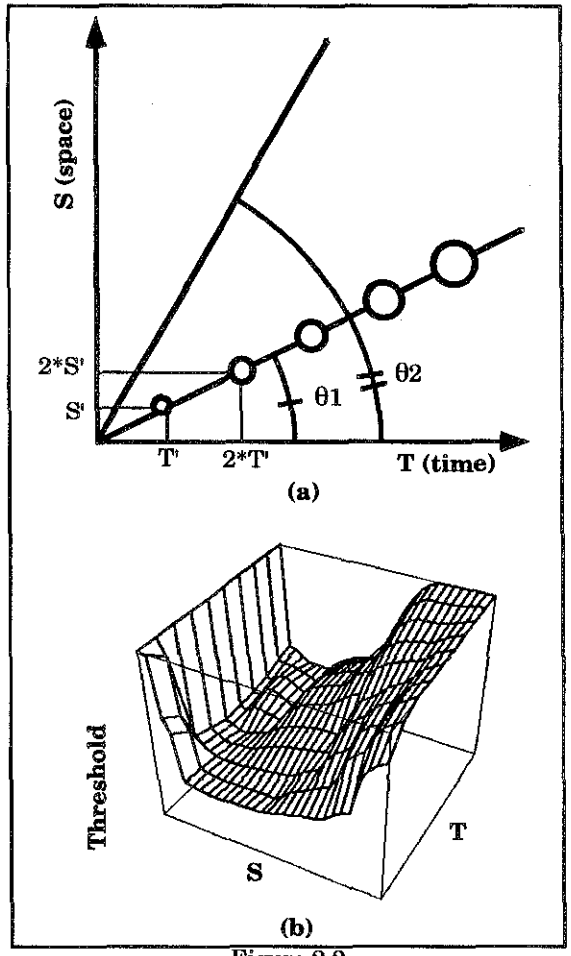


Figure 2.2

(a) The array of bi-local detectors in figure 2.1 can be mapped onto the <S,T> plane. The abscissa represents the differential temporal delay between the receptive fields and the ordinate represents spatial distance between the receptive field centers. A single detector maps onto a point in the plane. Speed is measured as the tangent of the angle between the line through the origin and the abscissa. A motion stimulus with $N_L > 1$ contains motion information at multiple points in the plane against a background of motion noise. (b) An example of a threshold surface over the <S,T> plane.

The stimulus will affect detectors in some region around each point, where the region size will depend on the spatio-temporal tuning bandwidths of the neighboring detectors (indicated by the circular activation areas in figure 2.2a). There is evidence that parameters of the bi-local detectors scale with E (van de Grind et al., 1986), and tuning bandwidth is likely to scale in a similar manner, so the total area in the <S,T> plane representing stimulated detectors is a function of N_D, N_L, S_S, T_S, the density distribution of detectors and the spatio-temporal tuning characteristics of the detectors. The logical stimulus for preferentially stimulating a single point in the <S,T> plane is a single displacement life time, multiple displacement motion stimulus ($N_L = 1$, $N_D > 1$).

Motion stimuli with $N_L < N_{L\text{-max}}$ inherently contain a certain amount of spatial noise because some dots must be randomly displaced on a periodic basis. Spatial signal-to-noise ratio (SSNR) can be defined as the ratio of the number of coherently displaced dots to the number of incoherently displaced dots. When $N_L = 1$, 50% of the dots are incoherently displaced between any two frames, resulting in SSNR = 1.0. The presence of spatial noise in a motion stimulus with $N_L = 1$ and $N_D > 1$ means that there is motion information at a large number of directions and magnitudes (the spatial noise) but with a predominant amount of information at a single $<S_S,T_S>$ value. The correlation properties of the bi-local detectors and the higher level computations that follow make it possible to extract the coherent motion information from the noisy background.

2.3 Threshold Determination

Directional motion perception threshold measurements taken over a range of $<S_S,T_S>$ values for a motion stimulus with $N_L = 1$ and $N_D > 1$ determine a threshold surface over the <S,T> plane (e.g. figure 2.2b)[1]. One method for determining a perceptual threshold for this stimulus is through modulation of SSNR by changing the ratio of coherently displaced dots

[1] There is evidence that motion perception thresholds vary with the inverse square root of the size of the stimulus (van Doorn, and Koenderink, 1984), as would follow from signal detection theory if independent detectors add probabilistically. From a modeling point of view it is simplest to assume that all of our hypothetical motion detectors are equally sensitive. Such an assumption provides a connection from our abstract model to our experimental data in that the sensitivity of the system to a stimulus with a single $<S_S,T_S>$ value provides a first order measure of the number of independent detectors in the system tuned to $<S_S,T_S>$. Information about model detector tuning bandwidths and sensitivities would be required for refinement of the model.

(signal) to incoherently displaced dots (noise). An alternative is what we will call the luminance signal-to-noise ratio (LSNR) method (van Doorn and Koenderink, 1982a, 1982b). In this method the luminance of each dot in the signal pattern is modulated between frames through pixel by pixel luminance addition of a spatially and temporally uncorrelated random noise pattern.

Power spectral density (PSD) analysis of each method (see appendix B) reveals that both methods have the characteristic of a flat PSD. However the SSNR method changes both the mean luminance (d.c. component of the Fourier transform) and the expected power at each non-zero spatial frequency (a.c. components of the Fourier transform) simultaneously for both the signal and the noise patterns. The LSNR method changes only the average power of the a.c. component of the signal and noise patterns. Because detectors of large and small S_S have different sized receptive fields [van de Grind et al, 1986] and receptive field size contributes to luminance response characteristics, changing the mean luminance may affect the threshold measurement in different ways for different S_S sizes. It is important to vary only one parameter of the stimulus to measure a threshold, so we have used the LSNR method. Note that, although our stimulus has a dot density of 50% and a 50% SSNR (because $N_L = 1$), these values are kept constant throughout the experiment. The threshold is then derived using the LSNR method.

We note here that the use of a stimulus with $N_L = 1$ and $N_D > 1$ is not a new concept (e.g. Snowden and Braddick, 1989a) nor is the use of the LSNR threshold method (van Doorn and Koenderink, 1982a, 1982b). The novelty of our approach comes from the combination of these two techniques and a constant stimulus duration rather than a constant N_D (to ensure saturation of threshold level for all T_S values). The result is a method that provides a bridge between traditional two frame ($N_D = 1$, $N_L = 1$) apparent motion results and continuous ($N_D >> 1$, $N_L = N_{L\text{-max}}$) apparent motion results.

2.4 Methods

2.4.1 Stimulus Generation

The motion stimulus was generated using custom image generation hardware driven by a Macintosh IIfx computer. The monitor had a P4 phosphor and a base display rate of 90 Hz. Stimulus frame times were integer multiples of the base display rate. The inter stimulus interval (ISI) was zero seconds in all experiments. The display screen was 14 centimeters and 256 pixels square, and pixels were 0.55 mm in diameter. Foveal experiments were performed at a distance of 4.0 meters, so the screen subtended 2.0 degrees of arc and each pixel subtended 0.47 minutes of arc. The threshold surface was also measured at non-zero eccentricities of 12 and 24 degrees. The stimulus size was scaled in the periphery according to the stimulus magnification $M_S = (1+E/E_2)$, where E is stimulus eccentricity in the visual field (in degrees), and E_2 is the visual field eccentricity where the property being scaled is doubled ($M_S = 2$). The previously determined E_2 value of 4 degrees for detector span scaling (van de Grind, van Doorn and Koenderink, 1983) was used. Thus at 12 degrees stimulus and pixel size were scaled by a factor of 4, and at 24 degrees they were scaled by a factor of 7. This scaling was achieved by changing the distance between the display and the subject. A black fixation dot (approximately 2 minutes of arc in width) was affixed to the center of the stimulus display area during the foveal experiments and a red light emitting diode was used for fixation during experiments in the periphery. The apparent motion stimulus used in all experiments reported here consisted of moving random dot patterns with pixel life times of two frames (one spatio-temporal displacement). Luminance modulation of the signal pattern by a noise pattern was used to measure a directional motion discrimination threshold for the signal pattern using a staircase procedure. The average luminance of the display was set to 50 cd-m^{-2}. The signal-plus-noise random dot pattern had an average contrast level of 70%, a value well above the motion threshold saturation value of 35% (van de Grind, Koenderink and van Doorn, 1987). The LSNR method has been previously described (van Doorn and Koenderink, 1982a, 1982b). The luminance of each pixel in the target pattern is independently modulated (luminance addition) by a random dot noise pattern. The noise pattern is completely renewed for each frame and is spatially uncorrelated. The noise

distribution is therefore both spatially independent (uncorrelated across the image at any given time) and temporally independent (uncorrelated at each pixel across time).

2.4.2 Threshold Measurement

All experiments were two alternative forced choice left-right motion discrimination so as to distinguish a direction discrimination threshold from a simple motion, flicker, or contrast change detection threshold. Left-right motion was used to discount anisotropies in motion direction discrimination ability (Scobey and Kan, 1991; van de Grind, Koenderink, van Doorn, Milders and Voerman, in press). Each stimulus presentation consisted of one second of stimulus display plus a short contrast fade-in/fade-out on a mean luminance background. The one second duration was informally verified to be an upper bound on temporal integration in central vision by having subject EF perform the same experiment using a 2.0 second stimulus duration in central vision. The 2.0 second stimulus duration had no effect on the shape of the threshold surface. Informal measurements of temporal integration properties (subject EF at 12 degrees) indicate that 1.0 second of stimulus presentation time is also adequate to provide saturation of response at the non-zero **E** values tested here. The thresholds were determined using a staircase procedure that tracked the 79% correct level; three consecutive correct directional discriminations resulted in a lowering of the LSNR by 2 dB, while any other sequence raised the LSNR by 2 dB. There were either 9 or 10 turning points in all experiments with the threshold calculated as the average of the final six turning points. Turning point sequences were inspected and any sequence that did not reach a stable fixed point due to attentional variation was rejected and re-measured. In order to decrease the required experimental time, the subjects were asked to manipulate the LSNR manually (before beginning each staircase procedure) until it was just above subjective threshold. If a subject mis-estimated the starting LSNR, it was apparent in the turning point sequence of the staircase procedure as a continuing upward trend without a stable fixed point. Such sequences were also rejected and the point re-measured.

The threshold surface was measured by holding a constant $\mathbf{T_s}$ and measuring the LSNR threshold for a range of retinal displacement magnitudes ($\mathbf{S_s}$). A single session consisted of up to 10 different $\mathbf{S_s}$ values. Subjects were allowed to rest after any or all of the individual $\langle \mathbf{S_s}, \mathbf{T_s} \rangle$ point threshold measurements. The experiments were performed in a canvas tent with ambient lighting provided by the stimulus display screen. All experiments were monocular, and the subjects used a chin and head rest. Subjects were instructed to fixate on the appropriate location and to determine the direction of motion of the stimulus. The subject's answer was recorded via a key press on the computer keyboard and the LSNR was automatically changed by the computer. All subjects had normal or corrected to normal vision.

2.5 Results

The stimulus was considered as non-perceivable if the threshold rose above 700 during the staircase procedure. Such points were plotted in the graphs at an LSNR value of 1000. Some minimum $\mathbf{S_d}$ values were not measurable with the equipment configuration, and in these cases only an upper bound was obtained. All threshold curves measured in these experiments are quantitatively and qualitatively reproducible. Higher variance for the portions of the threshold curve at higher LSNR values is at least in part due to the use of a multiplicative (rather than additive) step size in the staircase procedure.

Because the results presented here were obtained with a new paradigm and embrace a wider range of parameters than those previously reported, we must use care in defining the terms used to compare them to results obtained using other paradigms. The popular terms D_{min} and D_{max} (minimum and maximum perceivable spatial displacement), for example, contain no reference to a dependence on $\mathbf{T_s}$. The motion detectors responding to $\mathbf{S_s}$ represent the limits of available $\mathbf{S_d}$ combined with the fact that the motion detectors will respond to a range of $\mathbf{S_s}$ values.

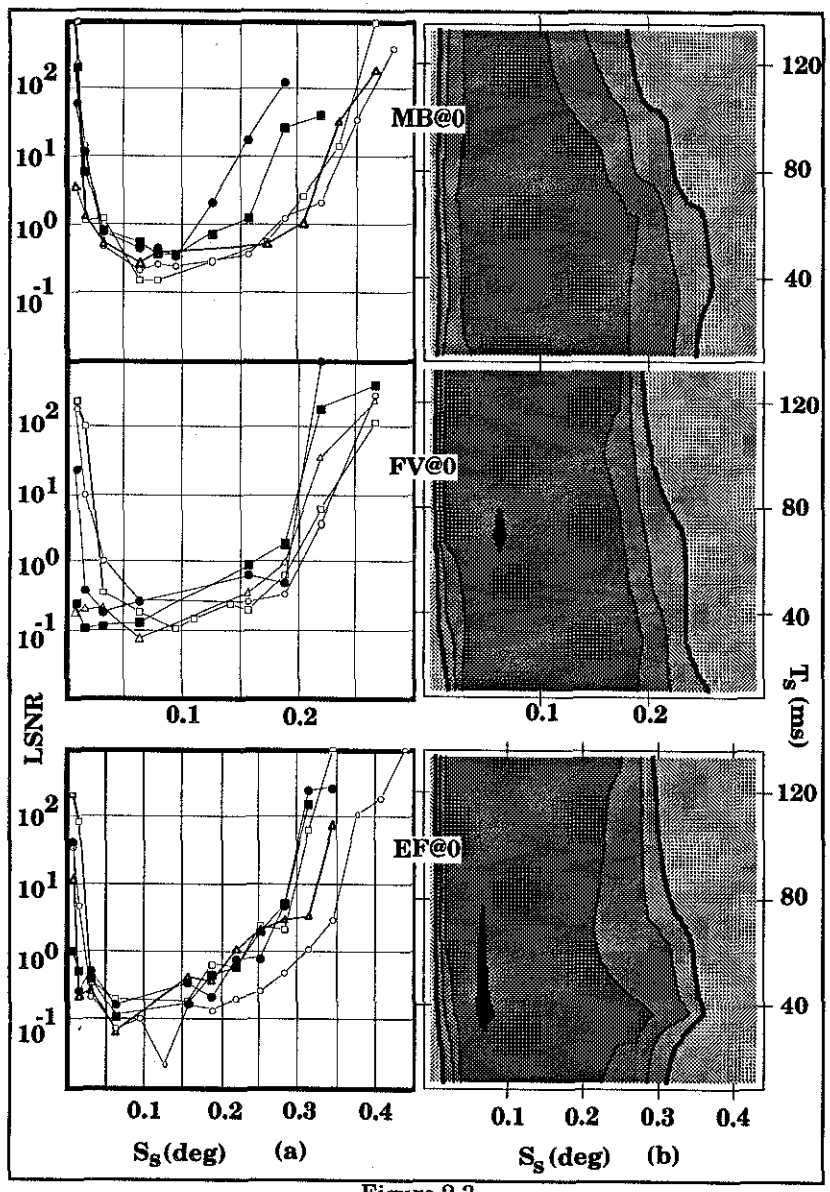


Figure 2.3

Figure 2.3a shows the raw threshold data at the fovea for three subjects plotted with S_S on the abscissa (in degrees), LSNR on the ordinate, and T_S as the parameter in each figure. □ = 11 ms, ○ = 33 ms, Δ = 66 ms, ■ = 100 ms and ● = 133 ms. Figure 2.3b shows the raw data from figure 2.3a re-plotted as a contour diagram of the corresponding threshold surface. Contour lines are drawn for LSNR values of 0.1, 1, 10 and 100, with the LSNR = 100 contour drawn in a thicker pen line. The enclosed regions are shaded from low to high LSNR as dark to light.

We do not have corresponding detector tuning bandwidth information in these experiments, so we will couch discussion in terms of S_S. Threshold values above an LSNR of 100 are considered as just visible because the surface is very steep at these places (e.g. see figure 2.3). That contour on the $<S_S,T_S>$ plane is therefore less insensitive to small changes in the surface shape and position. We will refer to the perceivable spatial displacement limits defined by the threshold contour LSNR = 100 as S^{100}. This contour is implicitly a function of T_S. The minimum perceivable S_S will be denoted by S^{100}_{min}, and the maximum perceivable S_S by S^{100}_{max}. This contour level is explicitly shown on the contour diagrams shown in the figures but does not represent all of the information represented by the surface. The term sensitivity is more general, but S^{100}_{min} and S^{100}_{max} must be used in comparison with previous work. We define sensitivity (equation 1) as inversely proportional to the measured LSNR threshold so that we may use the term that is clearest in a given context.

$$\text{Sensitivity} = \frac{1}{\text{LSNR}} \tag{2.1}$$

Figure 2.3a shows the raw data in central vision for each subject. The graphs show the threshold data plotted against S_S with T_S as a parameter. Figure 2.3b shows the same data but re-plotted as a contour diagram for each surface represented by the raw data. The contours are marked at each decade of LSNR {100,10,1,0.1}. The S^{100} contours in figure 2.3b are emphasized by a thicker line. The surface was created by re-sampling the raw data after linear interpolation of the points. Regions in the re-sampled area but outside the original sampling region were assigned the LSNR of 1000. Each contour diagram is shown on the same spatial and temporal scale as the original data. The re-sampling resulted in a low-pass filtering of the original surface. That filtering reduces the influence of measurement error but includes a consequent loss of information at

smaller S_S values. Low S_S regions of the surfaces are therefore reproduced on a different scale in figure 2.6 and are discussed later[2].

The central vision threshold surfaces have similar qualitative shapes for all three subjects. All show steeper slopes at the small S_S side of the surface than at the large S_S side of the surface. Subject FV shows a roughly constant S^{100}_{max} for the entire range of T_S values, with a slightly increased S^{100}_{max} for the lower T_S values. Subject EF also shows a roughly constant S^{100}_{max} for all T_S values, but with a more pronounced S^{100}_{max} optimum. Subject MB shows roughly S^{100}_{max} constancy for T_S values up to 66 ms but then shows a decreasing S^{100}_{max}. Beyond 133 ms S^{100}_{max} begins to decrease also for EF and FV (not shown), but this effect may be attributable to the change in the number of displacements that can be seen during the 1 second presentation time. At 133 ms the presentation time allows only about $N_D = 6$, and N_D has been shown to affect the displacement threshold for motion (Nakayama & Silverman, 1984; Snowden and Braddick, 1989a). Under some conditions $N_D = 6$ is the saturation limit for improvement of motion sensitivity (Snowden and Braddick, 1989b).

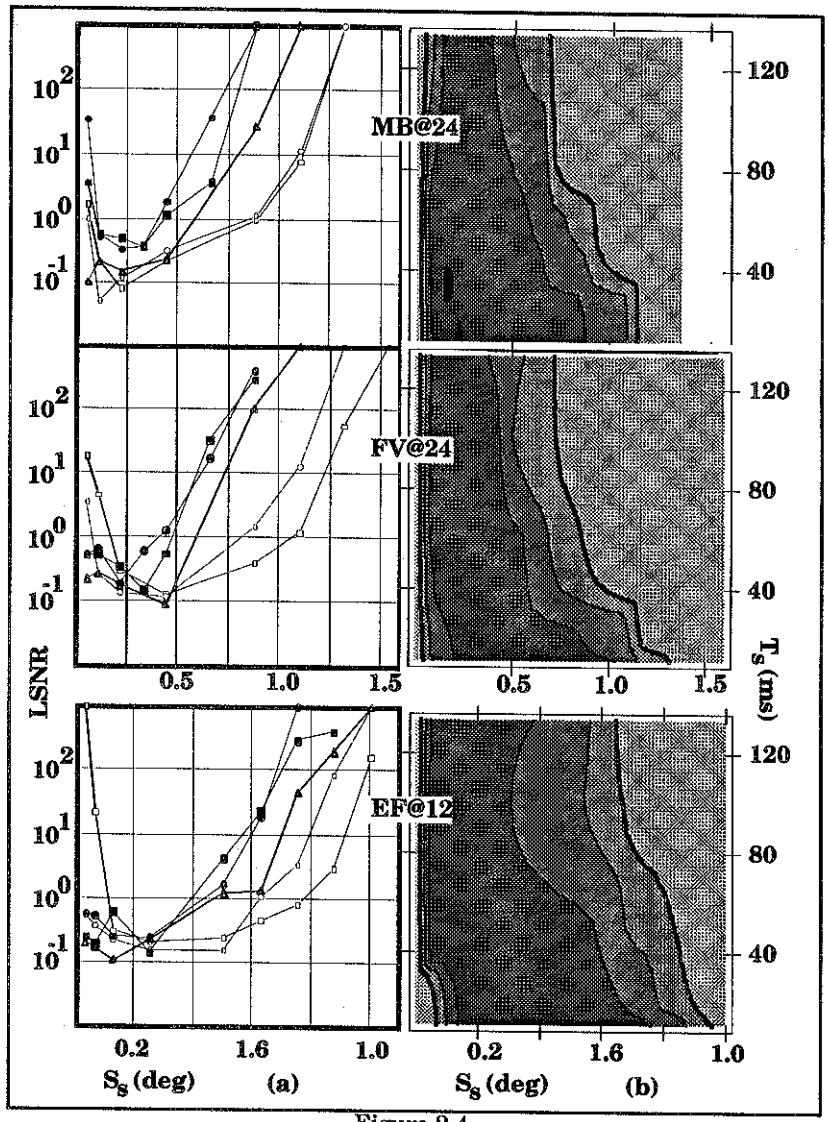


Figure 2.4
(a) The raw threshold data in the peripheral visual field for all subjects is plotted as in figure 2.3a. Subject EF performed the experiments at 12° from the fovea while subjects MB and FV performed them at 24°. (b) The raw threshold data from figure 5 are re-plotted as in figure 2.3b.

[2] Often a logarithmic scale can be used to display variations that are proportional to a variables value but in this case a logarithmic scale only suppressed the important variations at large S_S. We therefore chose to display the two regions separately.

For subjects EF and FV the constancy of S^{100}_{max} on the foveal threshold surface lasts into the region of 100 ms, or about $N_D = 9$. The threshold surface of subject MB, however, is constant only out to 66 ms (the third threshold curve in figure 2.3a), or about $N_D = 14$. Experiments controlling the stimulus for N_D rather than presentation duration (to be reported elsewhere) indicate that subject MB has a higher N_D saturation threshold (i.e. MB's sensitivity to the motion stimulus increases more slowly with displacement count) than either of the other two subjects. This corresponds with the earlier fall off in S^{100}_{max} for MB's foveal threshold surface. Note that although S^{100}_{max} is roughly constant centrally it still exhibits a dependency on T_S, moving to larger values at very short T_S values for subjects EF and MB.

Figure 2.4a shows the raw threshold data measured in the peripheral visual field for three subjects. The data are presented in the same format as figure 2.3a. Figure 2.4b shows the re-plotted data from figure 2.4a in the same manner as in figure 2.3b. The surfaces were measured at 12 degrees for subject EF and at 24 degrees for subjects MB and FV. The surfaces are very similar to each other in qualitative shape but are different from the foveal threshold surfaces. Again, all show steeper slopes on the surfaces at the small S_S side of the surface than at the large S_S side of the surface. S^{100}_{max} begins to fall off immediately with increasing T_S but then begins to stabilize around 66 ms (the third threshold curve in figure 2.4a). It is apparent that the use of $E_2 = 4$ degrees did not result in a perfect scaling of the maximum span properties. Furthermore, scaling by a simple constant is not appropriate as the peripheral surfaces do not have the same shape as the foveal surfaces. Instead it appears that the scaling of S^{100}_{max} must be dependent on T_S as well as E. If we use S^{100}_{max} as a fiducial point, we can calculate the value of E_2 required at each T_S for a correct scaling of perception of large S_S. The values calculated in this fashion are plotted in figure 2.5. There is no doubt that S^{100}_{max} is changing systematically with E, and that it is changing in a similar manner for each subject with T_S.

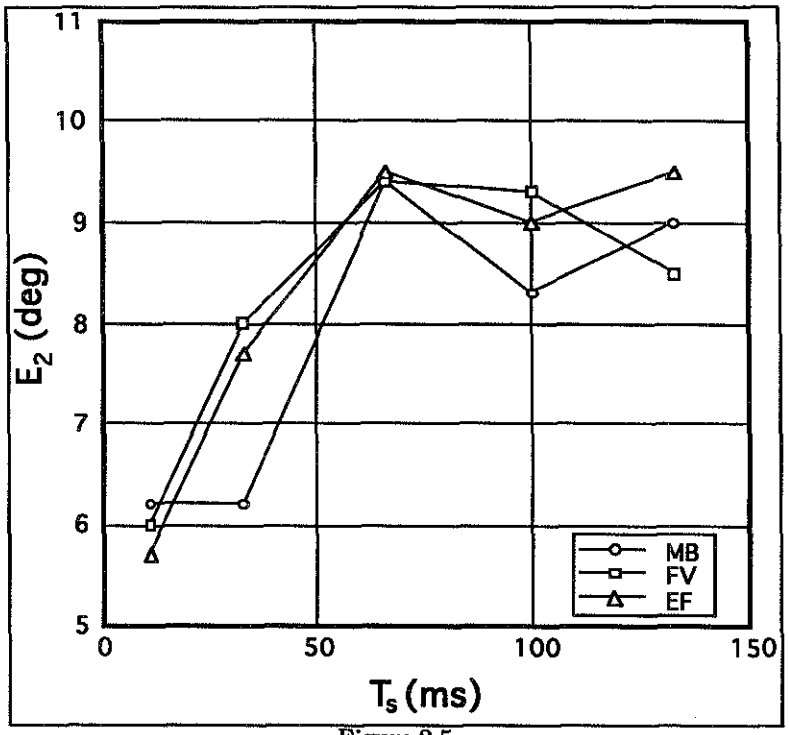


Figure 2.5
The E_2 values calculated from the threshold data across eccentricity are shown with T_S on the abscissa and E_2 on the ordinate. The contour LSNR = 100 was the reference point used at each value of T_S. The calculations for subject EF were performed for 0° and 12° while the values for subjects MB and FV were performed for 0° and 24°.

At non-zero E some of the S^{100}_{min} values around $T_S = 66$ ms were not measurable. This is not unexpected, as minimum perceivable S_S sizes have been shown to scale differently than the maximum perceivable S_S sizes (Snowden and Braddick, 1990). Subjects EF and FV re-measured the lower S_S range in peripheral vision using an E_2 value of 10.0 for determination of the stimulus size scaling factor.

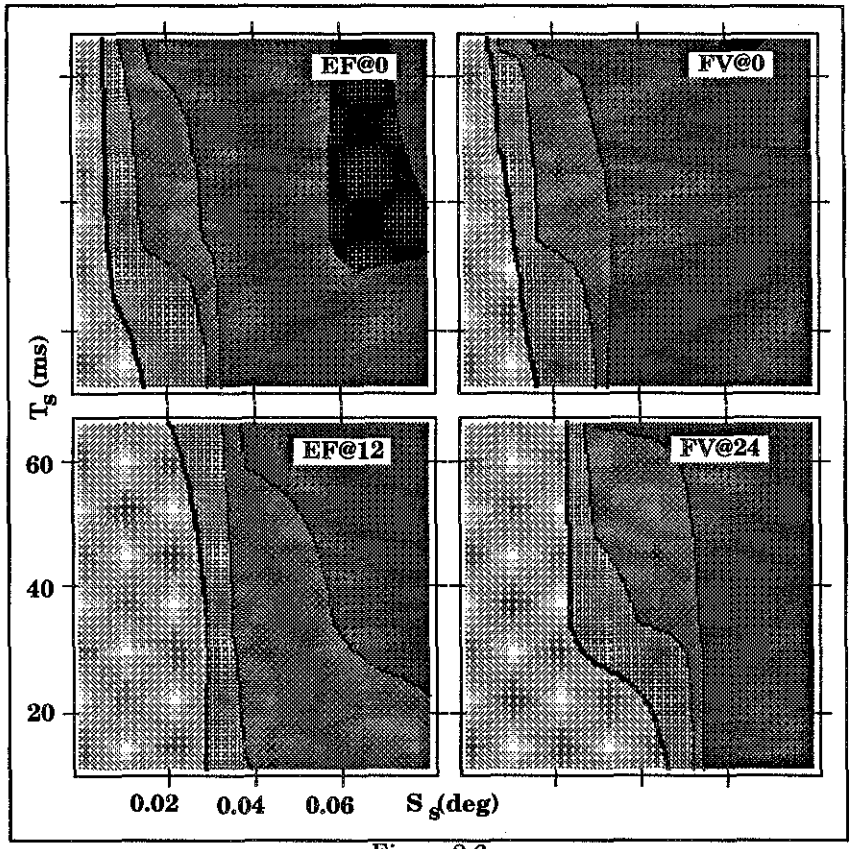


Figure 2.6
The thresholds for small S_S were re-measured for subjects EF and FV at 12° and 24°, respectively. The low <S,T> portions of the threshold surfaces for subjects EF and FV in central and peripheral vision are shown here on a larger scale in the form of a contour diagram. Contour lines are drawn for LSNR values of 0.1, 1, 10 and 100, with the LSNR = 100 contour drawn in a thicker pen line. The enclosed regions are shaded from low to high LSNR as dark to light.

This value of E_2 appeared to scale the stimulus more appropriately for small S_S, but we must be careful when considering such scaling because it includes the influence of spatial summation of the information in the stimulus as well as the extent of the stimulus on the retina, both of which influence the S^{100}_{min} and S^{100}_{max} curves. Figure 2.6 shows the contour diagrams of the low <S,T> portions of the four threshold surfaces. In all four diagrams it is obvious that the sensitivity to small S_S is better for T_S of 60-70 ms than for 11 ms (or for 133 ms, comparing with figure 2.4b with the spatial summation caveat mentioned above). Thus it seems that the S^{100}_{max} curve shows an optimal T_S value and a magnitude of dependence that increases with E.

2.6 Discussion

Previous experiments have been restricted either in the range of S_S, T_S, N_L and N_D combinations that could be tested (due to stimulus and/or threshold measurement method) or by the $<S_S,T_S>$ content of the stimulus (isolation of points in the <S,T> plane). Our experimental method has allowed us to more completely characterize the distribution of perceivable $<S_S,T_S>$ values (and indirectly, the distribution of S_d and T_d parameters on the human retina) than has previously been possible. These results can be related to a number of previous experiments.

Morgan and Ward (1980) used a stimulus paradigm in some ways similar to that used here to measure the perceivable range of S_S values in central vision for a range of interstimulus interval (ISI) values. They employed a single displacement life time random dot motion stimulus with a binary criterion for visibility under two different background lighting conditions. Individual dot presentation duration was very short (on the order of 1 ms), and dots were generated sequentially. The dot density is not reported, and the random dots were continuously updated (there were no frame boundaries, per se) so determination of an average pattern contrast is not simple. There is evidence that for short duration frame presentation times there is an upper ISI limit on the perception of directional motion (Baker and Braddick, 1985b). This means that their results for small ISI (delay in their figures) are comparable with those presented here. Even with widely different display conditions their thresholds for max and min perceivable S_S in central vision with variation of ISI are qualitatively similar to those reported here in the T_S region of 10-80 ms. Their results extend down to an equivalent ISI value of 2 ms, a delay that could not be reached with our equipment. The region of larger ISI in their results shows a cut off of perceivable spatial displacements at about 130 ms, a phenomenon also seen in later work (Baker and Braddick, 1985b). The results reported by

Morgan and Ward also indicate that background lighting can greatly influence the perceivable range of <S_S, ISI> values. High background lighting slightly increases the ability to perceive motion at small values of <S_S, ISI> while greatly impairing perception at large values of <S_S, ISI>, as compared to low background lighting conditions. Thus our choice of darkened viewing conditions appears to be the appropriate one if a single lighting condition is to be employed. Perhaps most importantly, it appears that very similar results can be obtained over a range of display conditions, indicating that this motion stimulus provides robust measurements.

D_{max} and D_{min} values have been measured at several eccentricities (Baker and Braddick, 1985a) using random dot patterns with $N_D = 1$ (and thus $N_L = 1$). Our findings agree with those results with respect to individual thresholds as well as the variation of thresholds across individuals. However, Baker and Braddick's (1985a) measurements of the change of D_{max} with exposure duration suggest that at non-zero **E** and short duration exposures (5-20 ms, their figure 7), D_{max} is smaller than for longer duration exposures. Our results indicate the opposite conclusion: in the periphery the largest perceivable displacements are seen with the smallest T_S values (of those used in this study). This difference is probably attributable to the use of $N_D = 1$ in the earlier experiments while the experiments reported here attempted to ensure a saturated detector response with respect to N_D. Because Baker and Braddick's (1985) results show that an ISI of 5 ms does not appreciably change their D_{max} (their figure 6) the addition of a 5 ms ISI should not appreciably change the shape of our threshold surface. It would seem then that the sensitivity of detectors to N_D is dependent on T_S, at least in the periphery. Interpretation of the two sets of experiments might lead to the prediction that the <S,T> threshold surface at non-zero **E**, with $N_D = 1$, would not show the characteristic dependence on T_S that we have seen here. Instead it would be roughly equivalent to that seen at the fovea for $N_D = 1$.

Measurements of D_{max} have also been made with ISI varied from 5 to more than 80 ms and exposure duration held constant at 60 ms (Baker and Braddick, 1985a). The resulting interpretation that increasing ISI decreases S_{max}^{100} is not directly comparable with our results because we used an ISI of zero in all experiments.

Our results can also be compared to some of those obtained using drifting 1 cycle-per-degree sinusoidal gratings with stimulus duration as an experimental parameter (Boulton, 1987). These experiments indicated that (1) at the fovea the minimum perceivable S_S is a constant, (2) in peripheral vision the minimum perceivable S_S follows a constant velocity, and (3), there is a smooth change between these two conditions from foveal to peripheral vision. A minimum perceivable velocity was measured in those experiments and used to calculate a minimum perceivable S_S. The measurement paradigm neither isolated points on the <S,T> plane nor controlled for temporal summation (N_D in our results[3]), so it is difficult to interpret in light of our current model. The quantitative values obtained with drifting sinusoidal gratings roughly correspond to those obtained for T_S values away from the global optimum region of 66-100 ms, but the present results indicate that minimum perceivable S_S is dependent on Ts at the fovea. The dependence of minimum perceivable S_S on a constant velocity might be interpretable in the following manner. S_{min}^{100} reaches a global minimum on the <S,T> plane in the region of 66-100 ms (our results), and this global minimum may remain on a line of constant velocity (a line through the origin of the <S,T> plane) regardless of stimulus duration. More precise interpretations will require a better understanding of how the limited spatial structure/frequency content of the grating image limits the stimulation of the bi-local detector population.

As the current data show, the independence of S_{min}^{100} and S_{max}^{100} from T_S depends on the scale at which we consider the contour behavior. At all **E** values measured here, S_{min}^{100} decreases as T_S increases from 11 ms to 66-100 ms. This dependence on T_S is qualitatively the same at the fovea as in peripheral vision with a (probably) smooth transition from lesser to greater

[3] The relationship between stimulus duration, T_S, and N_D in apparent motion may or may not be related as Duration = $(N_D+1)*T_S$. Such a relationship can be investigated through control of N_D in the stimulus.

dependence. The finding that dependence of S_{max}^{100} on T_S does not change with E (Baker and Braddick, 1985a) is also at odds with our results. The experiments presented here indicate that although S_{min}^{100} and S_{max}^{100} vary in a different manner across the visual field, that difference is only quantitative. S_{min}^{100} and S_{max}^{100} both decrease with increasing T_S (out to 66-100 ms for min S^{100}) at all values of E measured here. S_{min}^{100} then begins to increase again (as would be expected if detectors have a temporal tuning bandwidth) beyond 100 ms. These similarities are evidence that S_{min}^{100} and S_{max}^{100} (indeed the entire range of perceivable S_S) may be implemented by a single consistent mechanism rather than two different mechanisms (Snowden and Braddick, 1990).

This new information may also be interpreted in terms of previous experiments using continuous motion stimuli. Our results corroborate several of those of van de Grind et. al. (1986). Their experiments were designed to measure the monocular distribution of several properties of the bi-local detector population in humans using variable width random dot patterns. Bar width thresholds (i.e., a minimum motion-display-region width) were measured for speeds of mD/nT (D and T in their paper are the same stimulus parameters represented by S_S and T_S here) for various n and m values, using continuous motion ($N_L = N_{L\text{-}max}$) with and without the presence of LSNR. In terms of the current model, that method stimulated multiple points on the <S,T> plane. Because the motion was continuous rather than limited life time, the method of reducing bar width altered multiple variables simultaneously: spatial summation, the effective N_L, and SSNR (see appendix A). Interestingly, these experiments yielded results consistent with those reported here and elsewhere, probably in part because the reduction of bar width reduced the effective N_L. Previous to performing the experiments presented here subjects EF and FV repeated experiment 1 from van de Grind et al. (1986) in central vision. This experiment measured the largest stimulus stripe width (their parameter P1, figure 1) that just prevented perception of a horizontally moving motion stimulus. We found that if the measurements of these stimulus widths are used as an estimate of minimum span (consistent with the model presented here) for that velocity, then the method yields values within a factor of 3-4 of those reported here at low velocities (small S_S), and of 1-2 at higher velocities (large S_S). Reduction of the bar width such that $N_L = 1$ was apparently not possible due to loss of motion information in the stimulus at low velocities, while the increased SSNR at high velocities apparently also reduced the visibility of the motion. Their interpretation of a minimum span and a minimum delay are also supported by the results presented here. Thus the objection that the results of this experimental paradigm are affected by spatio-temporal recruitment area (Snowden and Braddick, 1989b) appears to be reasonable but does not affect their qualitative results and conclusions when the data are re-examined and interpreted within the current model framework.

2.7 Biological Interpretations

The shape of the <S,T> threshold surfaces presented here must be explicable in terms of neuroanatomy, neurophysiology, and signal processing theory. We propose the basic elements for a model based on the results of the experiments presented here and on well known and agreed upon properties of biological visual systems. The model is described without attempting to pin the individual components down to some preferred location or physiological implementation in the visual system. We will use the term scale interchangeably with spatial frequency, where large scale corresponds with low spatial frequency and small scale corresponds with high spatial frequency. We begin by assuming that the distribution of spatial scales of the receptive fields of the detectors changes with E in a manner at least similar to the change of receptive field sizes of, say, the primate retinal ganglion or dLGN cells. We also assume that a cell serving as a receptive field component in our model has spatio-temporal characteristics correlated with the receptive field spatial scale. Cells with large scale receptive fields tend to respond preferentially to larger scale image components and do so on a smaller temporal scale (a shorter, more transient temporal response), while cells with smaller scale receptive fields tend to respond to smaller scale image components and do so on a larger temporal scale (a longer, more sustained temporal response).

It is not unreasonable to postulate that if these components were to make up a system formed from a self-organizational process driven by available

visual input, similar components of the system would tend to group together in forming motion detectors. Such detectors are likely to reflect the spatio-temporal scale properties of the components making them up. That is, the set of detectors would have a distribution of spatio-temporal scale characteristics governed by the original components and the self-organizational process. There is evidence that detectors with larger scale receptive fields subserve larger span values (van de Grind, et. al., 1986; Boulton and Baker, 1991). This corresponds with the reasonable conclusion that the tuning bandwidth of a bi-local detector is a function of receptive field scale and the distance between receptive field centers. The result is a system that reflects a distribution of spatio-temporal response parameters $\mathbf{S_d}$ and $\mathbf{T_d}$ and their associated tuning characteristics. We next assume that the retinal receptive field locations of these components are governed by the cortical magnification factor: the mapping of the retinal ganglion cell projections, via the dLGN, to striate cortex. Our final assumption is that $\mathbf{S_d}$ is functionally implemented by a cortical neuronal interconnection distance distribution that is roughly constant across the region of cortex subserving the basic motion detection machinery.

The formation of the distributions of parameter values resulting from the self-organizational process of the visual system is beyond the scope of this paper. However there is one conclusion based on this sketch of a model that falls out of the model formulation. Any position in the visual field $\mathbf{E}$ has a density distribution of $\mathbf{S_d}$ values, and thus a range of perceivable $\mathbf{S_s}$ values, that is a result of following the mapping of visual space to cortex by a neuronal interconnection distances distribution that determines possible information correlation distances. Such a model results in a mapping (see appendix C) of distance $\mathbf{D_c}$ in visual cortex (in millimeters) to $\mathbf{S_s}$ values in the visual field by

$$S_s = (E_2 + E)\cdot\left|e^{\left(\frac{D_c}{M_f\cdot E_2}\right)} - 1\right| \quad \text{(deg)} \tag{2.2}$$

where E_2 and $\mathbf{E}$ are as described previously, and M_f is the cortical magnification at the fovea. S^{100}_{min} and S^{100}_{max} in this model would represent limits of correlational distances in visual space implemented by the limits of interconnection distances in cortex. Such a mapping determines a priori that S^{100}_{min} will be scaled differently than S^{100}_{max} and that the entire range of perceivable $\mathbf{S_s}$ values can be implemented by a uniform architecture. This non-linear mapping would also explain why the results of the experiments reported here do not clearly exhibit details of the threshold surface for both large and small $\mathbf{S_s}$ when displayed using a logarithmic $\mathbf{S}$ scale. Correct re-scaling would require use of (2) to scale all curves to a common cortical distribution. Previous work has also indicated that when a line is fitted to D_{min} and D_{max} data the D_{min} regression line has a non-zero intercept while the D_{max} regression line does not (Baker and Braddick, 1985a). Equation (2) indicates that the slope of the D_{max} regression line is large compared to that of D_{min}. The experimental error in the measurement procedure would therefore result in much larger relative error in the estimation of the D_{max} regression line intercept than for the D_{min} regression line intercept.

We must keep in mind that this is only a first order model of the distribution of perceivable pattern displacement sizes on the retina. The mathematical description does not account for variance of E_2 with $\mathbf{T_s}$, nor does it attempt to include correlation between receptive field size, spatio-temporal tuning, and span in the postulated bi-local detector (van de Grind, et. al., 1986). (These properties will be discussed in terms of self-organizational simulations presented in chapter 6.) Neither does it attempt to capture the multitude of neurophysiological and neuroanatomical details that influence the gradients of psychophysical thresholds in the peripheral visual field (Drasdo, 1991). However it does appear to provide a reasonable method for capturing the behavior of the perception of displacements of patterns on the retina.

2.8 Conclusions

We have presented a technique for exploring motion perception using a stimulus containing a predominance of isolated spatio-temporal displacement information. Our measurement of perceivable $\mathbf{S_s}$ and $\mathbf{T_s}$ at

several values of **E** using this stimulus indicate that motion perception paradigms using only one or a small number of displacements do not provide for optimum directional motion sensitivity for small $\mathbf{T_S}$ values. The results also show that S_{min}^{100} and S_{max}^{100} both depend on $\mathbf{T_S}$, and that this dependence changes with **E** in a qualitatively similar manner for both extrema, a result contrary to previous research. A description of likely biological properties of the bi-local detector array was presented that explains how detection of a wide range of $\mathbf{S_S}$ values can be performed using a homogenous system architecture and why such a system predicts that S_{min}^{100} and S_{max}^{100} scale with **E** in a qualitatively similar manner but by quantitatively different factors. The argument used by some researchers that this scaling difference is evidence that small and large displacements are perceived by different systems does not appear to be tenable. Further exploration of temporal (variation with $\mathbf{N_D}$) and spatial (variation with retinal coverage) summation in human motion perception using this new paradigm should provide the basis for a model that can be useful in interpreting the results of a wide variety of motion stimulus experiments in terms of the bi-local detector array model and in unifying the results of single (or limited) $\mathbf{N_D}$ paradigm experiments with those of continuous motion experiments.

Acknowledgments

Eric Fredericksen was supported by a UNC Cognitive Science Curriculum Pre-doctoral Fellowship, and an International Brain Research Organization Research Fellowship. Frans Verstraten is supported by the Biophysics Foundation of the Netherlands Organization for Scientific Research (NWO). We thank Marieta Braks for serving as an observer. The paper has been improved by the comments provided from critical readings of the manuscript by Jane Boulton and Andrea van Doorn, and through valuable discussions with O.-J. Grüsser. Special thanks go to Peter Schiphorst and Alexandra Bellekom for designing and building the custom stimulus generation hardware, and to Rob van Weerden for taking care of the 1001 peripheral details required for use of the equipment during experimentation.

Bibliography

Adelson, E. H. and Bergen, J. R (1985) Spatiotemporal energy models for the perception of motion. *JOSA A*, **2**, 284-299.

Baker, C. L. Jr. and Braddick, O. J. (1985a) Eccentricity-dependent scaling of the limits for short-range apparent motion perception. *Vision Research*, **25**, 803-812.

Baker, C. L. Jr. and Braddick, O. J. (1985b) Temporal properties of the short-range process in apparent motion. *Perception*, **14**, 181-192.

Ball, K. and Sekuler, R. (1982) A specific and enduring improvement in visual motion discrimination. *Science*, **218**, 697-698.

Ball, K. and Sekuler, R. (1979) Masking of motion by broadband and filtered directional noise. *Perception and Psychophysics*, **26**, 206-214.

Bischof, W. F. and Groner, M. (1985) Beyond the displacement limit: an analysis of short-range processes in apparent motion. *Vision Research*, **25**, 839-847.

Bischof, W. F. and Di Lollo, V. (1990) Perception of directional sampled motion in relation to displacement and spatial frequency: evidence of a unitary motion system. *Vision Research*, **30**, 1341-1362.

Boulton, J. C. (1987) Two mechanisms for the detection of slow motion. *JOSA A*, **4**, 1634-1642.

Boulton, J. C. and Baker, C. L. Jr. (1991) Motion detection is dependent on spatial frequency not size. *Vision Research*, **31**, 77-87.

Braddick, O. J. (1973) A short-range process in apparent motion. *Vision Research*, **14**, 519-527.

Browne, S. F. and McKee, S. P. (1989) Motion interference in speed discrimination. *JOSA A*, **6**, 1112-1121.

Chang, J. J. and Julesz, B. (1983) Displacement limits for spatial frequency filtered random-dot cinematograms in apparent motion. *Vision Research*, **23**, 1379-2385.

Drasdo, N., (1991) Neural substrates and threshold gradients of peripheral vision. In Vision and Visual Dysfunction, Vol. 5, Macmillan Press, pp. 251-265.

Grüsser, O.-J. (1991) Impairment of perception and recognition of face, facial expression and gestures in schizophrenic children and adolescents. In Eggers, C. (Ed.): Schizophrenia and youth, Springer, Heidelberg, 100-118.

Grüsser, O.-J., Landis, T. (1991) Visual agnosias and other disturbances of visual perception and cognition. In Vision and Visual Dysfunction, Vol. 12, MacMillan Press, London.

McKee, S. P. and Welch, L. (1985) Sequential recruitment in the discrimination of velocity. *JOSA A*, **2**, 243-251.

Morgan, M. J. and Ward, R. (1980) Conditions for motion flow in dynamic visual noise. *Vision Research*, **20**, pp. 431-435.

Nakayama, K. and Silverman, G. H. (1984) Temporal and spatial characteristics of the upper displacement limit for motion in random dots. *Vision Research*, **24**, 293-299.

Nakayama, K. and Tyler, C. W. (1981) Psychophysical isolation of movement sensitivity by removal of familiar position cues. *Vision Research*, **21**, 427-433.

Poggio, T. and Reichardt, W. (1973) Considerations on models of movement detection. *Kybernetik*, **13**, 223-227.

Reichardt, W. (1961) Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. In Sensory Communication, Rosenblith, ed. (Wiley, New York, 1961), pp. 303-317.

Santen, J. P. H. van and Sperling, G. (1985) Elaborated Reichardt detectors. *JOSA A*, **2**, 300-321.

Scobey, R. P. and Kan, P. L. E. (1991) A horizontal stripe of displacement sensitivity in the human visual field. *Vision Research*, **31**, 99-109.

Snowden, R. J. and Braddick, O. J. (1989a) Extension of displacement limits in multiple-exposure sequences of apparent motion. *Vision Research*, **29**, 1777-1787.

Snowden, R. J. and Braddick, O. J. (1989b) The combination of motion signals over time. *Vision Research*, **29**, 1621-1630.

Snowden, R. J. and Braddick, O. J. (1990) Differences in the processing of short-range apparent motion at small and large displacements. *Vision Research*, **30**, 1211-1222.

Snowden, R. J. and Braddick, O. J. (1991) The temporal integration and resolution of velocity signals. *Vision Research*, **5**, 907-914.

Stremler, F.G. (1982) Introduction to communication systems, 2nd Ed., Addison-Wesley Publishing Company, Inc., Reading, Massachusetts.

Van Doorn, A. J., and Koenderink, J. J. (1982a) Spatial properties of the visual detectability of moving spatial white noise. *Experimental Brain Research*, **45**, 189-198.

Van Doorn, A. J., and Koenderink, J. J. (1982b) Temporal properties of the visual detectability of moving spatial white noise. *Experimental Brain Research*, **45**, 179-188.

Van Doorn, A. J., and Koenderink, J. J. (1984) Spatiotemporal integration in the detection of coherent motion. *Vision Research*, **24**, 47-53.

Van de Grind, W. A., Van Doorn, A. J., and Koenderink, J. J. (1983) Detection of coherent movement in peripherally viewed random dot patterns. *JOSA*, **73**, 1674-1683.

Van de Grind, W. A., Koenderink, J. J. and Van Doorn, A. J., (1986) The distribution of human motion detector properties in th monocular visual field. *Vision Research*, **26**, 797-810.

Van de Grind, W. A., Koenderink, J. J., Van Doorn, A. J., Milders, M.V., and Voerman, H. (in press) Inhomogeneity and anisotropies for motion detection in the monocular visual field of human observers. To appear in *Vision Research*.

Van de Grind, W. A., Koenderink, J. J., and Van Doorn, A. J. (1987) Influence of contrast on foveal and peripheral detection of coherent motion in moving random-dot patterns. *JOSA A*, **4**, 1643-1652.

Williams, D. W. and Sekuler, R. (1984) Coherent global motion percepts from stochastic local motions. *Vision Research*, 24, 55-62.

Williams, W., Phillips, G. and Sekuler, R. (1986) Hysteresis in the perception of motion direction as evidence for neural cooperativity. *Nature*, **324**, 253-255.

Wilson, H. R., Levi, D., Maffei, L., Rovamo, J. and DeValois, R. (1990) The perception of form: Retina to striate cortex. In Visual Perception: The Neurophysiological Foundations, Spillman and Werner, eds. (Academic Press, Inc., San Diego, 1990) pp. 231-272.

Appendix A
Inherent Spatial Noise

A motion stimulus with $N_L = 1$ and $N_D > 1$ inherently contains a SSNR<=1.0 where the value depends on the displacement size of the stimulus. The SSNR for horizontal motion in a rectangular stimulus with $N_D = 1$ can be calculated as

$$\text{(A1)} \qquad SSNR = \frac{W - S_s}{W + S_s}$$

where **W** is the width of the screen and S_s is the size of the displacement. The equation is derived through consideration of the uncorrelated pixels that are placed onto the screen due to the spatial displacement and the initial 50% correlated dot content of the stimulus. For **W** = 256 pixels and S_s ranging from 1 to 40 pixels, SSNR in a motion stimulus with $N_L = 1$ goes from 0.99 to 0.73. How this affects directional perception threshold is undetermined, but the only alternative to keep the SSNR constant is to widen the display, which also changes the extent of the stimulus on the retina and therefore the available retinal span values in the detector array. This means that the minimum span values are likely to be overestimated and maximum span values are likely to be underestimated by some amount. This problem is inherent to all random dot motion displays, regardless of dot lifetime. For a continuous motion stimulus ($N_L = N_{L\text{-}max}$) we have

$$\text{(A2)} \qquad SSNR = \frac{W - S_s}{S_s}$$

where **W** is the width of the screen and S_s is the size of the displacement. Again, this is due to the addition of uncorrelated dots to the pattern as the spatial pattern is displaced. Equation (A2) decreases more slowly for a given screen width than equation (A1). In the work by van de Grind et al. (1986), **W** is a parameter that is varied so that it approaches the value of S_s, resulting in an appreciable increase in SSNR as defined by equation (A2)

Appendix B
Power Spectral Density of a Random Dot Pattern

The power spectral density (PSD) of a random signal can be calculated from the autocorrelation function of that signal (Stremler, 1982, pp. 471-477). For a discrete, uncorrelated random dot pattern that function has the form

$$\text{(B1)} \qquad \begin{array}{ll} A\delta(\tau) & \tau = 0 \\ D & \tau \neq 0 \end{array}$$

where d is the Dirac delta function and t is the correlation offset. The PSD is the Fourier transform of (1) which is

$$\text{(B2)} \qquad \begin{array}{ll} A & \omega \neq 0 \\ D\delta(\omega) & \omega = 0 \end{array}$$

where **A** is the average power of the a.c. components and **D** is the average power of the d.c. component of the spatial signal. This provides equal average power at all a.c. (spatial) frequencies superimposed on an average d.c. (mean luminance) power signal. Because the spatial patterns can be considered to be generated by random variables we can calculate **A** and **D** for any random dot pattern that is spatially uncorrelated and stationary (the statistics of the random variable do not change over time). For a random process (variable) X generating our discrete random dot pattern we have that (op. cit.)

$$\text{(B3)} \qquad D = \mu_X^2 = E[X]^2$$

$$\text{(B4)} \qquad A = \sigma_X^2 = E[X^2] - E[X]^2$$

where m and s are the mean and standard deviation of the variable. In linear systems theory the signal and noise characteristics can be considered separately although they may be added together, or superimposed. The stimulus patterns in the SSNR and LSNR threshold methods are such a case: two patterns, signal and noise, superimposed. What we wish to know is how the signal and noise characteristics change in the SSNR and LSNR

situations when a threshold is being measured. In the SSNR case the pattern (either signal or noise) has some dot density (ρ). This gives us

(B5) $D = \rho^2 L^2$

(B6) $A = \rho L^2 - \rho^2 L^2 = L^2(\rho - \rho^2)$

where ρ is the average dot density of the signal. In the LSNR case the pattern (signal or noise) has dot density ρ=50% with pixels of luminance $L+\Delta$ (dot present) and pixels of luminance $L-\Delta$ (dot not present). This gives us

(B7) $D = L^2$

(B8) $A = \Delta^2$

The two methods are complementary with respect to how the threshold is measured: the SSNR method holds the luminance values of the 1's and 0's constant while varying ρ, and the LSNR method holds ρ constant while varying the luminance values of the 1's and 0's. The difference in the two methods can be seen in how the values of **D** and **A** are affected. In the SSNR case the average power of the mean luminance and the average power of the spatial frequency components are varied simultaneously for both the signal and the noise patterns (altering the ratio A_{signal}/A_{noise} and the ratio $\mathbf{D_{signal}/D_{noise}}$). In the LSNR case the average power of the mean luminance is constant while the average power of the spatial frequency components is varied for both patterns. Since both luminance and spatial frequency content of a spatial pattern can affect the perception of that pattern, it is advantageous to use the LSNR threshold measurement method.

Appendix C

Mapping Derivation

The mapping of visual space onto visual cortex has been studied a great deal for both simple receptive field properties of retinal ganglion cells and their projection to visual cortex. The cortical magnification factor M, is often used as a simple model of the mapping of the visual field to visual cortex (Drasdo, 1991, pp. 251-265).

(C1) $$M = \frac{M_f}{\left(1 + \frac{E}{E_2}\right)} \quad \text{(mm/deg)}$$

$\mathbf{M_f}$ is the cortical magnification at the fovea, **E** is the eccentricity of the stimulus in the visual field, and $\mathbf{E_2}$ is the value of **E** at which the value of M is half that in the fovea. Further modifications of this basic formula are only for greater accuracy at larger **E** values and do not change the nature of the shape of the curve. Extension of this model to psychophysical measurements only requires new $\mathbf{M_f}$ and $\mathbf{E_2}$ values.

We are considering a situation where such a mapping is followed by a population of inter-neuronal interconnections (e.g. dendritic arborization and/or axonal projection) that mediates the correlation operation of a bi-local detector. Equation (1) above is just the instantaneous derivative of cortical distance $\mathbf{D_c}$ with respect to visual angle **E**. Integration of (1) with respect to visual angle (using the fovea as the origin) provides a formula for the distance in cortex representing a sweep of visual angle from the fovea to some point **E** on a radial line through the fovea (Wilson, Levi, Maffei, Rovamo and DeValois, 1990, pp. 235).

(C2) $$D_c = M_f \cdot E_2 \cdot \ln\left(1 + \frac{E}{E_2}\right) \quad \text{(mm)}$$

If we pick two points p1 and p2 on the radial line, the mapping of the distance between the two points p1 and p2 from visual space to visual cortex is just

(C3) $$D_c = M_f \cdot E_2 \cdot \ln\left(\frac{E_2 + E_{p2}}{E_2 + E_{p1}}\right) \quad \text{(mm)}$$

Setting p1 = **E**, and p2 = **E** + S_s gives us an approximate mapping of distance in visual cortex, D_c, to a corresponding spatial displacement S_s at the position E in the visual field.

(C4) $$D_c = M_f \cdot E_2 \cdot \ln\left(\frac{E_2 + E + S_s}{E_2 + E}\right) \quad \text{(mm)}$$

Inverting (4) gives the distance in S_s for a given combination of D_c and **E**.

(C5) $$S_s = (E_2 + E) \cdot \left(e^{\left(\frac{D_c}{M_f \cdot E_2}\right)} - 1 \right) \quad \text{(deg)}$$

This equation is linear in **E** and the slope can be radically different depending on D_c. The slopes of D_{min} and D_{max} with respect to **E** for subject CLB in (Baker and Braddick, 1985a) are (fitting the published data by hand) 0.002 degrees/degree (S_s/**E**) and 0.19 degrees/degree, respectively (a factor of 95).

In an independent line of research Grüsser (1991; see also Grüsser and Landis, 1991 p. 178) used a similar approach to develop a formula of exactly the same form as (5); only the interpretation of some of the constants might differ. Interestingly, he showed that the formula could describe the foveofugal motion of a migraine phosphene front across the field of view, by assuming that the corresponding brain disturbance moves at a constant cortical velocity (on the order of 2.2 mm/s).

Chapter 3*: The temporal integration of apparent motion

R. E. Fredericksen, F. A. J. Verstraten and W. A. van de Grind

Abstract

Human motion perception is assumed to be functionally described by an array of bi-local detectors feeding later, higher order computational stages. Using this model as a guide, improvement of spatio-temporal displacement sensitivity by temporal integration (summation) was measured in human central vision using random dot apparent motion stimuli. Our results agree with previous experiments with regard to improvement of maximum perceivable spatial displacement but show that contrary to previous results the minimum perceivable spatial displacement can be improved in a similar manner. Furthermore, it appears that stimulus duration is a better indicator of sensitivity improvement than the number of frames in the stimulus and that temporal tuning of motion detectors is inversely related to the size of the spatial pattern displacement.

3.1 Introduction

If we are present to see Sergey Bubka clear a new world record in the pole vault or to see Ellen van Langen come from behind to win the Olympic 800m, our eyes are receiving visual information about objects that move continuously through space. When we see these events recorded on our video tapes or films, however, we are viewing images that are snapshots of the original event. These images represent instantaneous, discrete displacements rather than continuous, smooth displacement of objects through space. This form of motion information, generally called apparent motion, has been used extensively in the study of visual motion perception. Such apparent motion information can have significantly different spatio-temporal structure than the information that we receive from real world events. Our visual systems evolved in the absence of such artificial motion information, but knowledge of how apparent motion is perceived can shed light on how our visual systems compute visual motion information.

The topic of interest in the psychophysical results presented here is how our perception of apparent motion is affected by stimulus duration. By stimulus duration we mean the total time that moving spatial patterns are viewed, that is, the total time that spatio-temporal information is available to the viewer. Stimulus duration in real motion is simply the total time that we view the scene. Stimulus duration in apparent motion, however, is a function of frame count (the number of snapshots in our sequence), individual frame duration (the amount of time each snapshot is visible) and interstimulus interval time (the time between snapshots when no spatial pattern is displayed). Each of these parameters can be varied throughout the snapshot sequence, and each can affect the visibility of the motion stimulus to the viewer. By visibility, we mean the strength of the motion signals extracted by the motion detectors in the viewer's visual system. This motion signal requires some amount of time to be calculated by the motion detector and increases in strength as the spatio-temporal information is integrated over time. The motion signal strength stops increasing when it reaches some maximum signal strength, a condition that we refer to as saturation of response in the rest of the paper. The equivalent behavior in terms of the measured threshold values is a flattening of the threshold curve (no further improvement of the threshold is possible).

The experiments presented in this paper were designed to investigate how the strength of the calculated motion signal increases over time and how much time is required for the motion signal to reach a maximum strength. Investigation of temporal integration of apparent motion can be complicated by the form of the image sequence and by the possibility of using various image frame rates and spatial displacement sizes. Moreover, a constant stimulus duration can be achieved using different

* The material in this chapter has been submitted to the journal *Vision Research* in almost identical form under the title "Temporal integration of random dot apparent motion information in human central vision" by the above authors.

combinations of frame rate and frame count. We therefore begin by describing our stimulus and the relationship among the parameters of frame rate, image frame count, and stimulus duration. We use stimuli with a constant image-frame exposure duration and no interstimulus interval. If we count the number of spatial displacements present in this type of apparent motion stimulus, stimulus duration can be calculated as

$$D_S(T_S, N_D) = T_S \cdot (N_D + 1) \tag{3.1}$$

where D_S is the total stimulus duration, N_D is the number of spatial displacements, or one less than the frame count , and T_S is the display time for each image, or the frame exposure duration. Some investigations of temporal integration in almost-true (i.e. very high frame rate) motion (Burr, 1981) and apparent motion (van Doorn and Koenderink, 1982b, 1984; McKee and Welch, 1985; Snowden and Braddick, 1991) have used D_S as the primary stimulus parameter. Other work on temporal integration and related properties in apparent motion (Nakayama and Silverman, 1984; Snowden and Braddick, 1989a, 1989b, 1990) has been reported in terms of N_D rather than D_S. Moreover it has been suggested that the temporal integration mechanism in apparent motion reaches a saturation point after a constant number of spatial displacements (Snowden and Braddick, 1989a). D_S and N_D are related by the value of T_S so choosing to investigating temporal integration in terms of N_D is a logical place to begin. However, we have recently shown that T_S can have a significant effect on the perception of the motion stimulus. Human directional motion perception thresholds measured using an apparent motion stimulus with D_S = 1 second are significantly different from those measured using a constant (small) N_D, and the threshold with D_S = 1 second is significantly dependent on the value of T_S for stimuli outside of central vision *(Fredericksen, Verstraten and van de Grind, in press; chapter 2). Because T_S influences the visibility of the apparent motion we cannot assume that the motion detector signal is independent of T_S. The experiments reported here were therefore designed to explore how D_S, T_S, and N_D each influence the motion signal calculated by the motion detectors in human central vision.

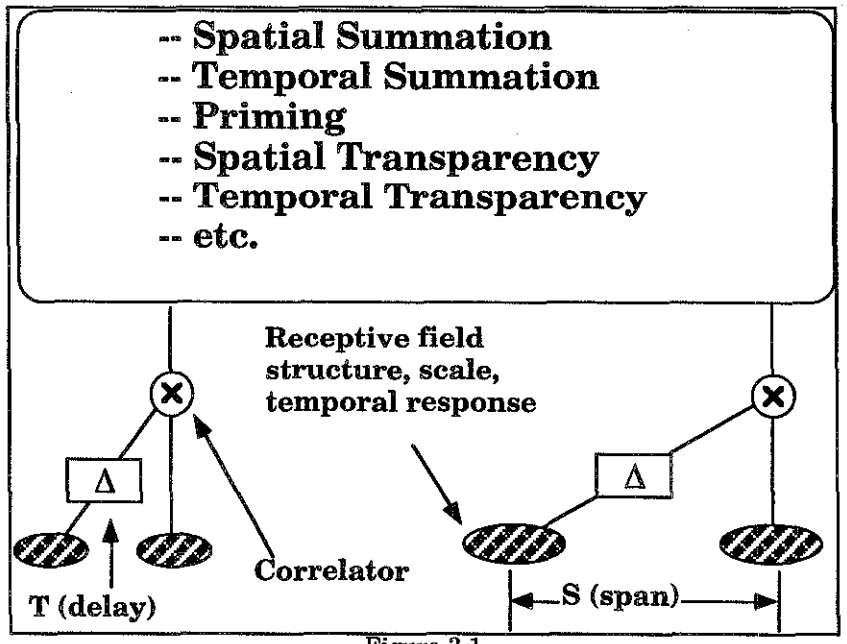


Figure 3.1
The functional performance of the human visual motion perception system is modeled as an array of bi-local detectors whose outputs are fed to later computational stages. The detector array can be characterized by distributions of detector span (distance between receptive field centers) and delay (differential delay of the receptive field components) at each position in the visual field. The receptive field components send their output to the correlation component of the detector, and can be characterized by their spatial structure, scale and temporal response.

In a previous paper we described an apparent motion stimulus designed to isolate specific populations of spatio-temporally tuned motion detectors (Fredericksen et. al., in press; chapter 2). The design assumed a functional description of human motion perception that is generally agreed upon: a front end array of correlational devices, in this case bi-local detectors (Reichardt, 1961), whose outputs are used in higher level computational stages (figure 3.1). This isolation is also important in our investigation of temporal integration because it can reveal whether detector sub-populations perform temporal integration differently. We

therefore employed the same class of stimulus in the experiments presented here.

In order to simplify discussion of the stimulus and its interaction with the model of motion perception used here, we briefly restate the model assumptions and stimulus selection criteria, and define terminology that will be used in the remainder of the paper. The bi-local detectors in the front end of the model shown in figure 3.1 are assumed to be characterized by the parameters depicted in the figure: (1) receptive field structure (sensitivity to spatial structure in the visual image), (2) distance between receptive field centers on the retina (span), and (3) differential temporal delay for the output of the receptive fields (delay). The motion information is generated through correlation of the differentially delayed activity of the receptive fields, and the information is fed into the higher level computations. We refer to the span and delay values of the detector as S_d and T_d, respectively. We use this model as our guide, but without further information we cannot determine which $<S_d,T_d>$ combinations are reacting to our stimulus, so we will refer to the corresponding stimulus parameters throughout the rest of the paper. Spatial displacement of the stimulus pattern and temporal delay (frame exposure duration) between spatial displacements of the stimulus pattern will be referred to as S_S and T_S, respectively. The reasons for using a random dot pattern in this situation are enumerated in Fredericksen et. al. (in press; chapter 2).

We also define a few terms to distinguish between the properties of unlimited-lifetime and limited-lifetime random dot pattern motion stimuli. We have already defined the number of spatial displacements in a stimulus as N_D, and the stimulus duration as $D_S = (N_D+1)*T_S$. In random dot apparent motion a dot is placed on the screen and is spatially displaced by an amount S_S, every T_S seconds (assuming no interstimulus interval) until it is no longer visible (its position falls outside the viewing rectangle). We can, however, limit any given dot to a small number of coherent spatial steps of size S_S before it is removed from the moving pattern. New random dots are then added in a manner consistent with the dot density and direction of motion. We refer to the number of displacements that a dot undergoes as *dot life time* (N_L). N_L controls the spatio-temporal displacement content of the stimulus and allows isolation of detector populations with specific $<S_S,T_S>$ tuning characteristics. Limiting dot life times to a single displacement (two image frames, or $N_L = 1$) removes any spatial correlation across image frames that are not adjacent in the image sequence and thereby allows preferential stimulation of detectors that respond to a single $<S_S,T_S>$ combination. For details see Fredericksen et. al. (in press; chapter 2). In these experiments a directional motion perception threshold was measured using a luminance signal-to-noise ratio (LSNR) paradigm. The luminance of the dots in the target random dot pattern is modulated through addition of the luminance of a corresponding spatially and temporally uncorrelated (noise) random dot pattern. Some reasons for use of such a paradigm have been enumerated in Fredericksen et al. (in press; chapter 2). The method was developed and has been described in detail elsewhere (van Doorn and Koenderink, 1982a, b).

The $N_D = 1$ stimulus paradigm was used to investigate the effect of temporal integration on the perceivability (measured in terms of LSNR) of a stimulus with various $<S_S,T_S,N_D>$ values. The experimental results presented here show that maximum and minimum perceivable S_S both depend on N_D and T_S, and do so in a similar manner. The data indicate that the value of N_D required for saturation of threshold (saturation of detector response to the motion information) depends on T_S, and that D_S is a better predictor of temporal integration than N_D. There is also evidence that the temporal tuning of motion detectors is inversely related to their spatial tuning.

3.2 Methods

3.2.1 Stimulus Generation

The motion stimuli were generated using custom image generation hardware driven by a Macintosh IIfx computer. The monitor had a P4 phosphor and a display rate of 90 Hz, or a base frame exposure time of about 11 ms. All T_S values used in these experiments were integer multiples of the base frame duration. The inter-frame interval was negligible in all experiments. The display screen was 14 centimeters and 256 pixels square, and pixels were 0.55 mm in diameter. All experiments were performed at a distance of 4.0 meters. The motion display was 2.0

degrees wide in all experiments with the display height reported in each experiment. Each pixel subtended 0.47 minutes of arc. A black fixation dot (approximately 2 minutes of arc in width) was affixed to the center of the stimulus display area. The apparent motion stimulus used in all experiments reported here consisted of moving random dot patterns with pixel life times of two frames (one spatio-temporal displacement). Luminance modulation of the signal pattern by a noise pattern was used to measure a perception threshold for the signal pattern using a staircase procedure. The average luminance of the display was set to 50 cd/m^2. The signal-plus-noise random dot pattern had an average contrast level of 70%, a value well above the motion contrast threshold saturation value of 35% (van de Grind, Koenderink and van Doorn, 1987). The luminance of each pixel in the target pattern is independently modulated by weighted addition with the luminance of the pixels of a corresponding random dot noise pattern that is spatially and temporally uncorrelated.

3.2.2 Threshold Measurement

All experiments were two alternative forced choice horizontal (leftwards or rightwards) motion discrimination so as to distinguish a direction discrimination threshold from a simple motion, flicker, or contrast change detection threshold. Horizontal motion was used to discount anisotropies in motion direction discrimination ability (Scobey and Kan, 1991; van de Grind, Koenderink, van Doorn, Milders and Voerman, in press). Each stimulus presentation consisted of a number of frames of motion superimposed on a mean luminance background. The thresholds were determined using a staircase procedure that tracked the 79% correct level; three consecutive correct directional discriminations resulted in a lowering of the LSNR, while any other sequence raised the LSNR. There were 10 turning points in all experiments with the threshold calculated as the average of the final six turning points. A step size of 2 dB was used at all times. Turning point sequences were inspected and any sequence that did not reach a stable fixed point due to attentional variation was rejected and re-measured. In order to decrease the required experimental time, the subjects were asked to manipulate the LSNR manually (before beginning each staircase procedure) until it was just above subjective threshold. If a subject mis-estimated the starting LSNR, it was apparent in the turning point sequence of the staircase procedure as a continuing upward trend in the turning points. Such sequences were also rejected and the point re-measured. Once threshold was reached, the turning point sequences were remarkably stable. Subjects were allowed to rest after any or all of the individual $\langle S_S, T_S \rangle$ point threshold measurements. The experiments were performed in an area screened off from stray light sources with ambient lighting provided by the stimulus display. All experiments were monocular, and the subjects used a chin and head rest. Subjects were instructed to fixate on the black dot on the screen and to determine the direction of motion in the stimulus. The subject's answer was recorded via a key press on the computer keyboard, and the LSNR was automatically changed by the computer.

3.2.3 Subjects

Three subjects participated in the experiments: 2 males (EF and FV) and one female (MB). All subjects were experienced observers and were the same subjects as in our previous publication (Fredericksen et. al., in press; chapter 2). All subjects had normal or corrected to normal vision.

3.3 Results

The motion was considered as non-detectable if the threshold rose above 700 during the staircase procedure. Points in the $\langle S_S, T_S, N_D \rangle$ parameter space that were non-perceivable were arbitrarily assigned an LSNR of 1000. Some minimum perceivable S_S values were not measurable with the equipment configuration, and in these cases only an upper bound was obtained. All threshold curves measured in these experiments are quantitatively and qualitatively reproducible. Threshold values above an LSNR of 100 are considered as just visible because the surface is very steep at these places (e.g. see figure 2.3). That contour on the $\langle S_S, T_S \rangle$ plane is therefore relatively insensitive to small changes in the surface shape and position. In order to simplify the comparison of our results with previous work, we define S^{100} as the threshold surface contour where LSNR = 100. We will refer to the contours defining the minimum and maximum perceivable S_S (both are functions of T_S) as S^{100}_{min} and S^{100}_{max}. To further simplify discussion, we define sensitivity as inversely

proportional to the measured LSNR threshold so that we may employ the term that is clearest in a given context.

3.3.1 Experiment 1 Description

Following the lead of previous research, this experiment was designed to investigate the dependence of human directional motion perception on N_D in the motion stimulus. Left-right directional motion detection threshold surfaces were measured using the LSNR method (as described above) for three observers over a range of $\langle S_S, T_S, N_D \rangle$ values. The display was 2 degrees in height and 2 degrees in width for all data taken in this experiment. The data describe a three dimensional surface in 4 dimensions, $\langle S_S, T_S, N_D, \text{LSNR} \rangle$, specified by up to 260 data points. The $\langle S_S, T_S, N_D \rangle$ values used in this experiment were chosen to ensure good sampling of the $\langle S_S, T_S \rangle$ subspace as determined in previous experiments (Fredericksen et. al., in press; chapter 2) and to cover some of the same N_D values chosen by previous experimenters.

3.3.2 Experiment 1 Results and Discussion

The raw data for only two observers are shown (subjects EF and FV) because of the large data volume (the raw data for the third subject is in the appendix). The data of the third subject (MB) are very similar both quantitatively and qualitatively. Figure 3.2 displays the raw data for subject EF. Each graph shows a set of LSNR threshold curves as a function of S_S. The parameter of the curves within each graph is N_D, and the parameter across the graphs is T_S. Figure 3.3 shows the data for subject FV in the same format. Initial inspection of these data (figures 3.2 and 3) indicate that subject EF is more sensitive to small S_S in the stimulus than subject FV. That is, subject EF's threshold for small S_S values improves more quickly with increasing N_D than does that of FV. The thresholds of subject MB increased least quickly of the three. The data in figure 3.2 for subject EF show that threshold curves change with N_D similarly for T_S values out to 133 ms, the largest T_S condition in a previous constant stimulus duration (D_S = 1.0) experiment (Fredericksen et. al., in press; chapter 2).

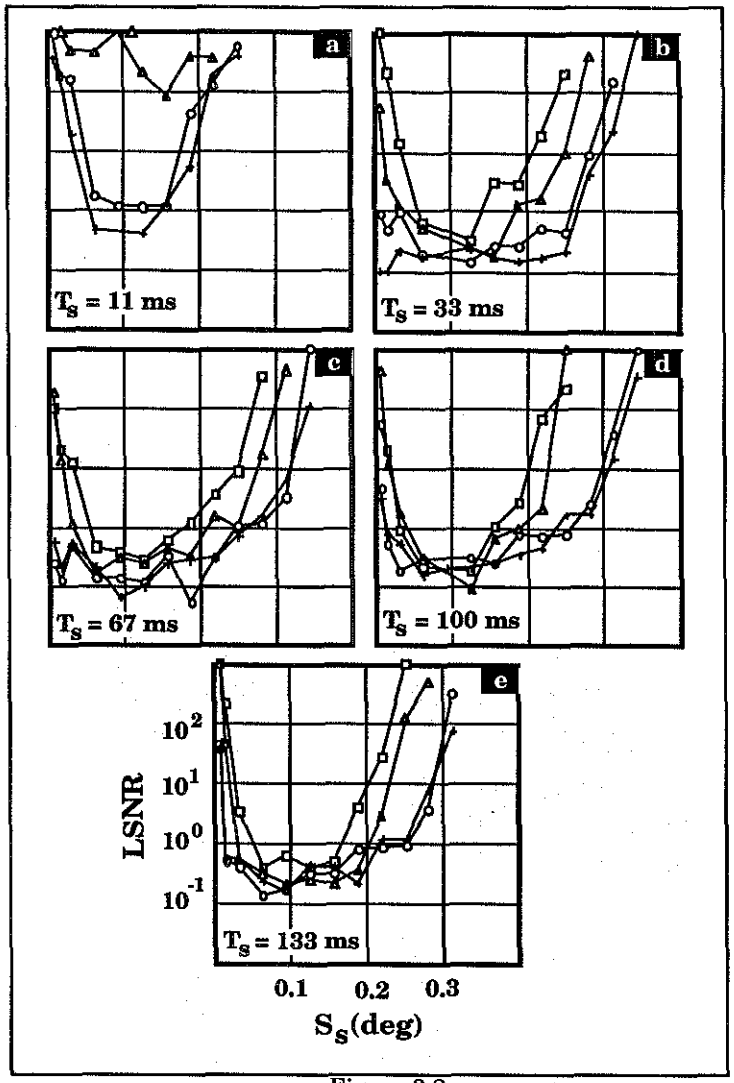


Figure 3.2
The raw data for subject EF is shown. Each graph displays a set of LSNR threshold curves as a function of S_S. The parameter of the curves within each graph is N_D. Unique symbols are assigned to represent a constant displacement count: □ for N_D=1, Δ for N_D=2, o for N_D=6, and + for N_D=10. The parameter across the graphs is T_S and each graph is labeled accordingly.

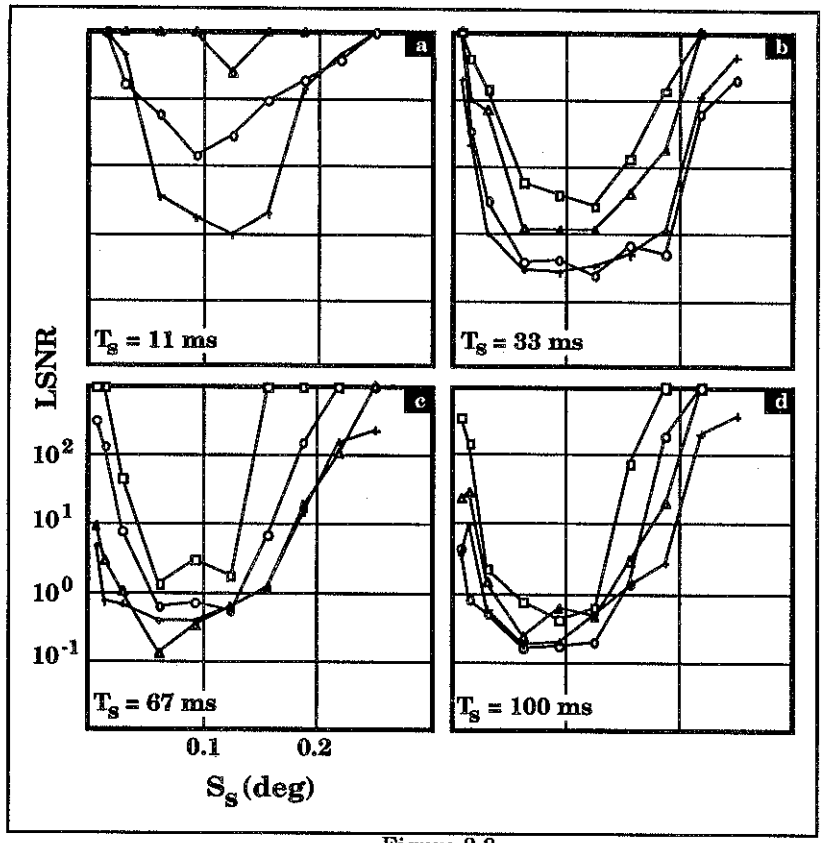


Figure 3.3
The raw data for subject FV is shown as in figure 3.2.

Figures 3.2 and 3.3 show the threshold data in a manner following the example of previous work, but that format is perhaps not the most illuminating. The data form a set of threshold measurements over a range of values for three variables: S_S, T_S, and N_D. Fixing any one of these three coordinates provides a "slice" through a four space that is a two dimensional threshold surface in three dimensions. Figure 3.4 shows the data of subject FV re-plotted as contour diagrams of the threshold surfaces obtained by fixing the value of N_D, while figure 3.5 shows contours of the surfaces obtained by fixing the value of T_S (the data for the other two subjects are shown in the same format in the appendix). Contours are shown every half log unit with the S^{100} contour drawn with a wider pen. Dark areas indicate low LSNR values (high sensitivity) while light areas indicate high LSNR values (low sensitivity). The figures show that dependence of the threshold on T_S is similar to the dependence of threshold on N_D. Sensitivity and thus S^{100}_{min} and S^{100}_{max} generally improves with both increasing N_D and increasing T_S, but some T_S dependencies appear in figure 3.4.

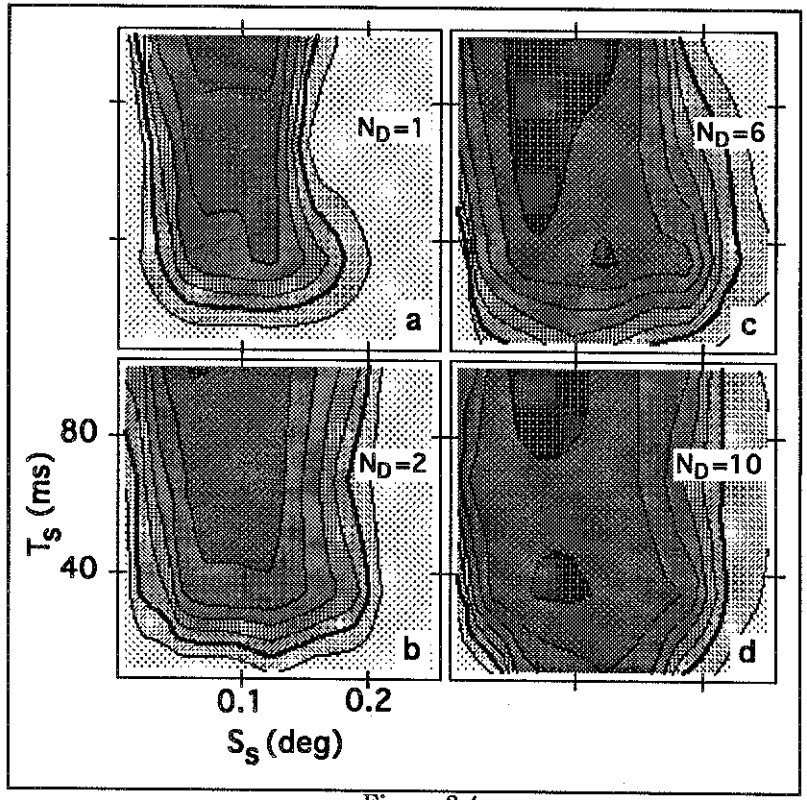


Figure 3.4
The raw data of subject FV are re-plotted as contour diagrams of the threshold surfaces obtained by fixing the value of N_D. Contours are shown every half log unit with the S^{100} contour drawn with a wider pen. Dark areas indicate low LSNR values (high sensitivity) while light areas indicate high LSNR values (low sensitivity).

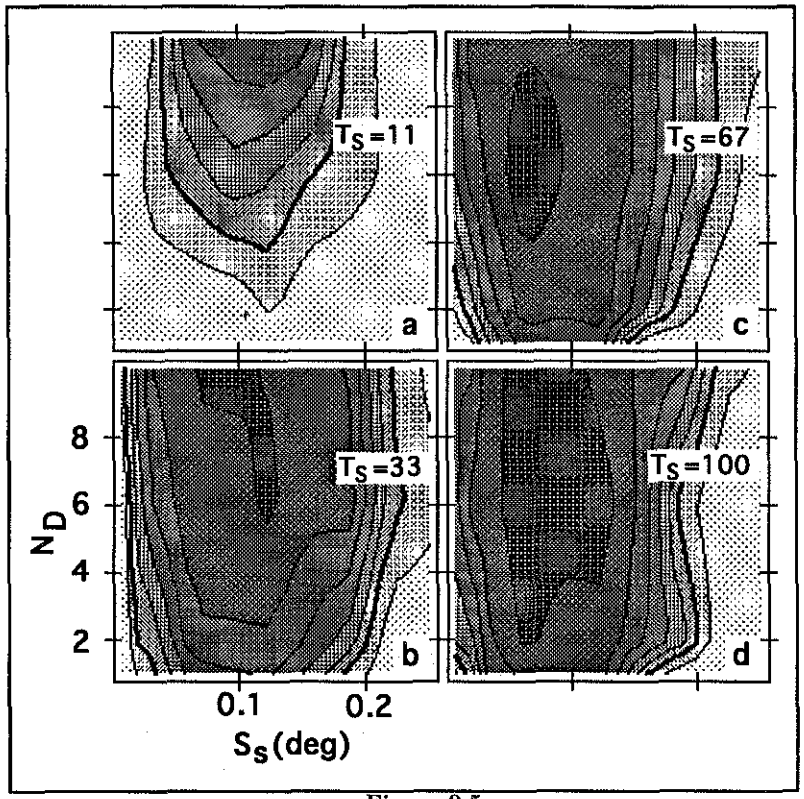


Figure 3.5
The raw data of subject FV re-plotted as contour diagrams of the threshold surfaces obtained by fixing the value of T_S.

Contours are shown every half log unit with the S^{100} contour drawn with a wider pen. Dark areas indicate low LSNR values (high sensitivity) while light areas indicate high LSNR values (low sensitivity).

Figure 3.4a (N_D = 1) shows evidence of dual S^{100}_{max} maxima: one at large T_S and one at small T_S in the $<S_S,T_S>$ surface. This effect is also seen for subject MB (figures in appendix) at N_D = 2 but is not seen for subject EF, probably due to a faster saturation of threshold. An even more interesting phenomenon in the $<S_S,T_S>$ surfaces is that the best S_S (by "best" we mean the S_S value to which the subject is most sensitive) for T_S = 11 ms begins above 0.1 degrees when N_D is held at 2 (figure 3.4b) and then moves to lower S_S values as N_D increases to 10 (figure 3.4d). The same effect holds for the $<S_S,N_D>$ surface as T_S is increased. The best S_S value at N_D = 1 begins above 0.1 degrees when T_S is held constant at 11 ms (figure 3.5a) and moves below 0.1 degrees as T_S is increased to 100 ms (figure 3.5d). This effect also appears in the data of subjects EF and MB (see appendix). An important difference between the behavior of the surfaces shown in figures 4 and 5 is that the surfaces for fixed T_S in figure 3.5 show essentially monotonic improvement of sensitivity with increasing N_D but the surfaces with fixed N_D do not show monotonic improvement with increasing T_S. The data of subject EF behaves in the same manner, but the behavior is not as clear for the data of subject MB. We hypothesize that this indicates the presence of T_S dependent tuning in the motion detector population. This tuning appears as non-monotonic S^{100}_{max} dependence on T_S within the individual graphs in figure 3.4 but across the individual graphs in figure 3.5. S^{100}_{max} in the graphs in figure 3.5 reaches its largest values (the whole curve moves to large S_S values) when T_S is in the range of 33 to 66 ms.

Increasing D_S in each graph in figures 3.4 and 3.5 is along the vertical axis. The similarity of behavior in the data across the N_D and T_S dimensions would indicate that D_S might be a better indicator of temporal integration if we account for T_S tuning. If we fix the value of S_S in our data set, we get surfaces over the $<T_S,N_D>$ plane. Figure 3.6a shows a contour diagram of a surface representing an estimate of what the constant-LSNR contours might look like if the perceptual threshold is inversely proportional to the total stimulus duration.

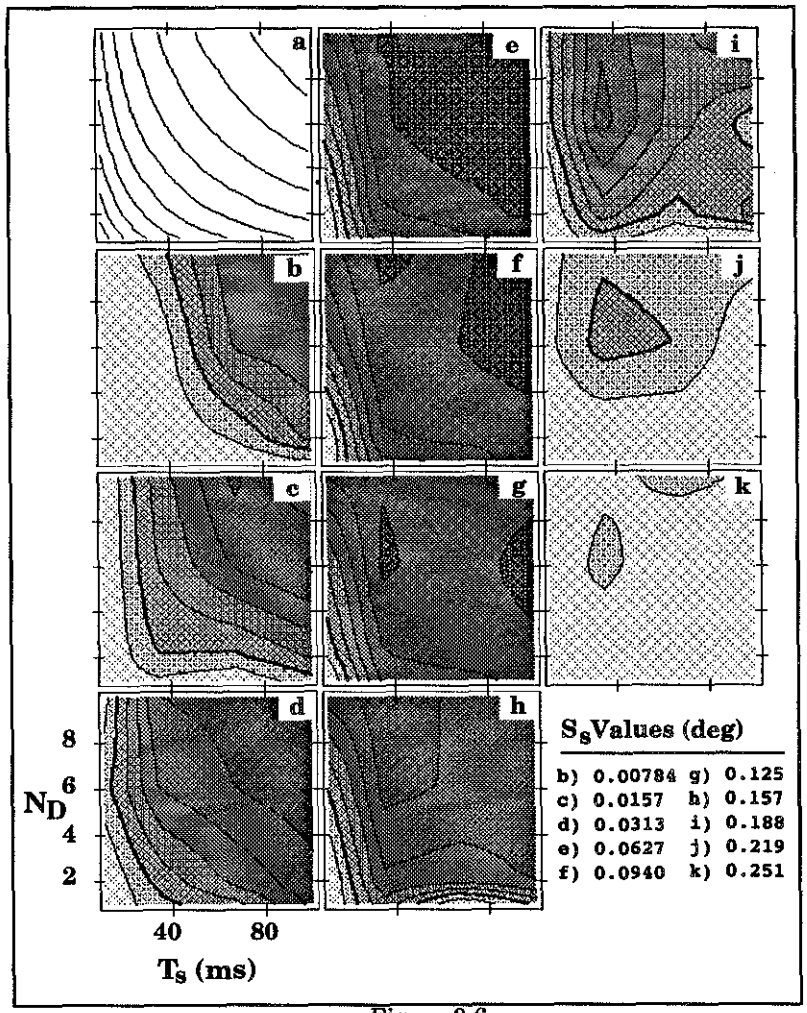


Figure 3.6

Figure 3.6a shows a contour diagram of a surface representing an estimate of what the constant-LSNR contours might look like if the perceptual threshold is inversely proportional to the total stimulus duration. Figures 3.6b-k show threshold surfaces for subject FV obtained by fixing the value of S_S. These LSNR threshold surfaces over the $<T_S,N_D>$ plane are plotted in the same manner as in figures 3.4 and 3.5. The legend included in the lower right hand corner of the figure indicates the fixed S_S value in each graph.

The model of temporal summation shown in figure 3.6a was formulated using just a few basic ideas about the operation of the visual system. A motion detector responding to an apparent motion stimulus fires at some average rate during the presentation of the stimulus. The average signaling rate is a function of the spatio-temporal energy collected by the motion detector, and the signal coming from the motion detecting neuron is available for temporal summation. Although the summation process may include thresholding and integrator response saturation effects, the simplest integration function is summation of that signal over time. The total stimulus duration as defined in equation 1 in the text is then our simplest measure of sensitivity improvement over time, and threshold values are inversely proportional to the sensitivity. The contour diagram in figure 3.6a is then defined as

$$\text{LSNR}_{est} = \frac{K}{D_S(T_S, N_D)} \tag{3.2}$$

for some constant K. To make the surface contours comparable to the measured threshold surfaces, we use a logarithmic LSNR-axis scale.

If D_S is a good measure of sensitivity, constant-LSNR contours over the $<T_S,N_D>$ plane in figures 3.6b-k should have a shape similar to the constant-D_S contours of the model in figure 3.6a. Figures 3.6b-k show threshold surfaces for FV over the $<T_S,N_D>$ plane plotted in the same manner as in figures 3.4 and 3.5 but obtained by fixing the value of S_S. Although the data sampling was designed primarily to illuminate the role of N_D rather than D_S, the similarity of the constant-LSNR contour shapes in figures 3.6b-h (the T_S tuning dominates at the largest S_S values in figures 3.6i-k) to the constant-D_S lines in figure 3.6a would seem to indicate that D_S is a better overall measure of threshold improvement than just N_D. The dual temporal tuning optima again appear in figures 3.6f and 3.6g. Figures 3.6e-g also show a saturation of threshold indicated by wider spacing of contours in the regions of large D_S. This saturation effect is highly pronounced in the data of subject EF, and similar to that of FV in the case of subject MB. Assuming that temporal contrast sensitivity influences temporal integration, sensitivity loss at high frame rates might appear as a decrease in contour spacing for

small T_S values independent of N_D. The data in figures 3.6e-h seem to show evidence of such a reduction of sensitivity for frame rates > 30 Hz (T_S < 33 ms). This view of the data also shows evidence that temporal tuning is inversely related to S_S. At small S_S (figure 3.6b) the surface shows a single sensitivity maximum for large T_S (for the range of $<T_S,N_D>$ values tested here). As S_S increases, the optimum sensitivity moves to smaller T_S values and tends toward a single optimum at the largest S_S values tested here (figures 3.6i-k). This property also appears in the data of subject MB. The lower T_S optimum appears to be continuously present for subject EF, but this observation is complicated by the subject's high sensitivity and resulting threshold saturation.

As shown in figures 4 and 5, both S^{100}_{min} and S^{100}_{max} move to more extreme values with increasing N_D, leading to a wider range of perceivable spatial displacements in human motion perception. However, the rate of improvement with N_D depends on T_S, and S^{100}_{min} shows most improvement when $N_D > 2$. Thus the experimental results confirm the previous finding that S^{100}_{max} improves with increasing N_D in a random dot pattern stimulus but also show that S^{100}_{min} can be improved as well if enough spatial displacements are included in the stimulus. It is also interesting that the threshold saturation time limits in the first experiment appear to depend on spatial displacement size (figure 3.6). If we choose the LSNR = 1 contour line (the fourth contour from the S^{100} contour, or the second lowest LSNR contour) as a convenient measure of the time required for threshold saturation, then figure 3.6 would indicate that the stimuli containing the middle range of S_S values measured here (figures 3.6e-g) reach threshold saturation faster than the stimuli containing the largest and smallest S_S values.

3.3.3 Experiment 2 Description

The second experiment was designed to test two additional properties of motion perception. The first is the threshold saturation effect hypothesized for subject EF (and to a lesser extent for subjects FV and MB in the middle S_S range) for the data measured in experiment 1. The second was to provide a better display of any temporal contrast sensitivity loss due to fast frame rates (90 Hz maximum in these experiments) by sampling the $<T_S,N_D>$ plane more frequently. The threshold saturation effect is tested by manipulation of spatial summation of the stimulus. Reducing the height of the stimulus (in degrees of visual angle) results in a loss of sensitivity that is proportional to the square root of the change in stimulus area, following probabilistic summation over space (van Doorn and Koenderink, 1984). Increasing the sampling frequency of the $<T_S,N_D>$ plane also provides a smoother threshold surface reconstruction that should remove sampling artifacts in the surfaces shown in figure 3.6. Experiment 2 was performed under the same conditions as experiment 1 with the following changes. Subject EF re-measured the $<T_S,N_D>$ threshold surface for S_S = 0.0313 degrees with the additional $<T_S,N_D>$ combinations of T_S = 22 ms and N_D = 4, and with various stimulus sizes. The stimulus sizes were (height x width) 2x2 degrees, 2x1 degrees, 2x0.5 degrees, and a fourth size in which the central 2x1 degree strip was removed leaving two, 2x0.5 degree strips above and below the fixation point. The portions of the 2x2 stimulus display area that did not contain random dot motion displayed the stimulus mean luminance value of 50 cd/m^2 without any random dots, either static or in motion (a flat luminance profile).

3.3.4 Experiment 2 Results and Discussion

Figure 3.7a shows the contour diagrams of the threshold surface at S_S = 0.0313 degrees for subject EF from the data in experiment 1. Figure 3.7b shows a theoretical surface like that in figure 3.6a but includes an estimate of the influence of temporal contrast sensitivity loss. In the physiological system temporal contrast sensitivity loss would represent a decreased average activity of the motion detecting neuron due to a decrease in total collected energy. If we think of the detector's activity as a current, temporal contrast sensitivity loss can be considered as a decrease in that current as the frame rate increases.

Classical contrast sensitivity experiments have employed stimuli such as flickering spots, and sinusoidal luminance gratings that either reversed the sign of the contrast over time or drifted in the visual field (e.g. see Van de Grind and Grüsser, 1973, or Kelly and Burbeck, 1984, for reviews

on the subject). Our stimulus can be considered as a combination of many of these classical contrast sensitivity stimuli, and there is no clear way to apply the extant information to estimate the quantitative effects of temporal contrast sensitivity loss in our experiments. We can, however, estimate the *shape* of temporal contrast sensitivity fall-off based on data from the literature. On a log-log scale (i.e., log(sensitivity) vs. log(temporal frequency)) the falloff of sensitivity from the maximum (ignoring local optima for some spatial frequencies) behaves similarly to an exponential. The function

$$CS(T_s) = 10^{-e^{(A\cdot\log(T_S)+B)}} \tag{3.4}$$

was chosen because it provides a smooth falloff of sensitivity as the frame rate increases. The constant B is included in order to provide a sharper knee in the curve (as compared to a simple exponential function) on the log-log plot. The values of A and B were selected (on the basis of temporal contrast sensitivity data for a pilot experiment using subject EF; data not shown) so that the function reached a value of approximately 1 at a 15 Hz frame rate and dropped to 0.5 at a 90 Hz frame rate. The estimated threshold surface in figure 3.7b is then

$$LSNR_{est} = \frac{K}{D(T_s, N_D)\cdot CS(T_s)} \tag{3.5}$$

Figures 3.7c-f show contour diagrams of threshold surfaces for subject EF (displayed in the same manner as figures 3.4-6) obtained by measuring additional samples at T_s = 22 ms and N_D = 4. The small black dots superimposed on the contour diagrams in figures 3.7a, and 3.7c-f indicate the points on the plane where threshold samples were taken. The display configuration is included as an icon in each figure.

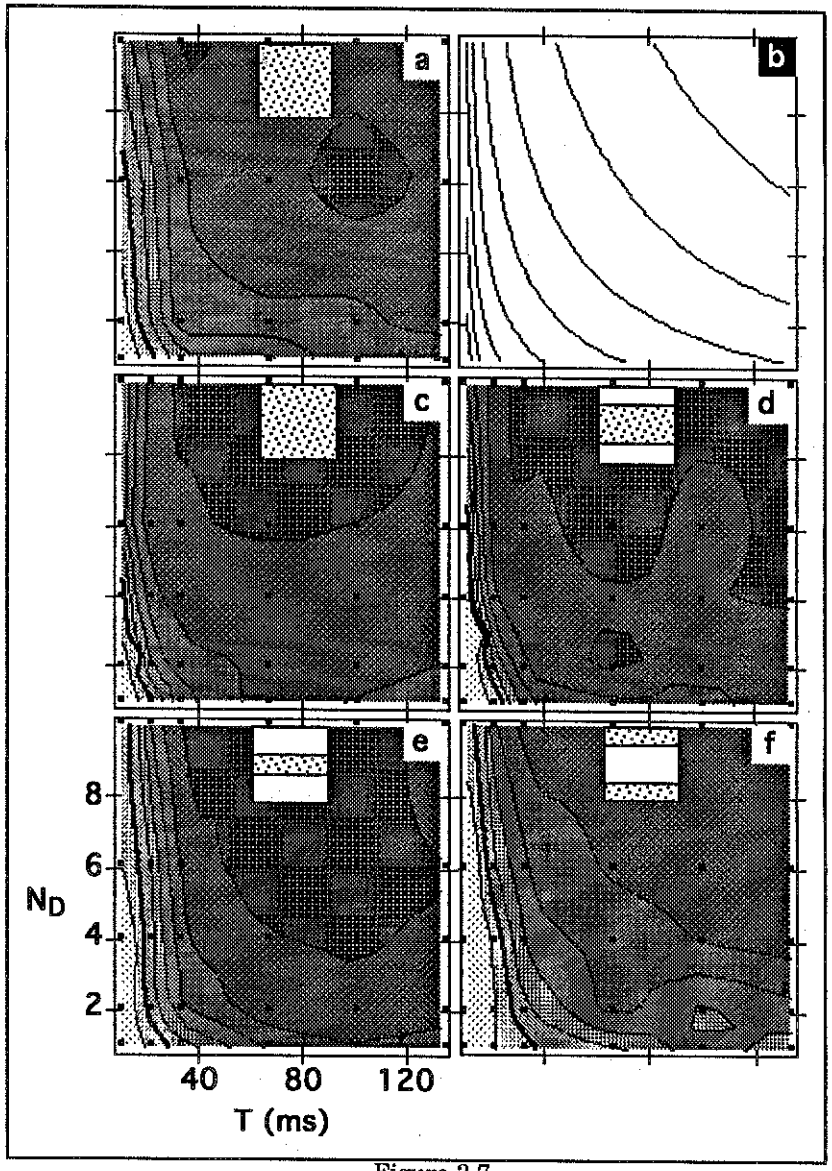


Figure 3.7
The figure shows the contour diagrams of the $<T_s,N_D>$ threshold surface at S_s = 0.0313 degrees for subject EF from the data in experiment 1. Figure 3.7b shows a theoretical surface like that in figure 6a but includes an estimate of the influence of temporal contrast sensitivity loss. Figures 3.7c-f show contour diagrams of threshold surfaces for subject EF

(displayed in the same manner as figures 4-6) on a sampling grid with additional samples at T_S = 22 ms and N_D = 4. The small black dots superimposed on the contour diagrams in figures 7a, and 7c-f indicate the points on the sampling points on the plane. The display configuration is included as an icon in each figure. The square icon represents the available stimulus display region with mean luminance regions represented by a white fill and motion stimulus regions represented by a dotted fill.

The surfaces in figures 3.7c-f show evidence of a decreasing amount of threshold saturation as the stimulus area is decreased. This can be seen by the increase spacing of the lowest LSNR contours and reduction of the low, flat area for large D_S values between figures 3.7c and 3.7f. For example, the LSNR = 1 contour (the fourth contour to the right of the S^{100} contour) moves out into the center of the plane as the display area is changed from the configuration in figure 3.7c to the configuration in figure 3.7f. We attribute this lowered amount of threshold saturation to the lowering of spatial summation resulting from the smaller display area. The deviation of the data in figures 3.7c-f from the calculated curves in figure 3.6b also appears to conform to that which would be expected from contrast sensitivity loss due to high frame rates as modeled in figure 3.7b.

The re-sampled surfaces again appear to be consistent with that expected if LSNR thresholds are a function of D_S. Interestingly, figure 3.7f shows the greatest similarity to figure 3.7b and greater overall reduction of threshold saturation effects than the surface in figure 7e, even though the threshold surface in figure 7e was obtained using one half of the total stimulus display area used in figure 3.7f. There may be several reasons for this. The very low (almost zero) luminance of the area surrounding the 2x2 degree stimulus, in simultaneous contrast with the upper and lower edges of the 50 cd/m^2 mean luminance stimulus area, may provide some sort of interference with the perception of the motion. However, since S_S in the stimulus was very small (0.0313 degrees, or about 1.9 minutes of arc) and there is evidence that receptive field size is proportional to the span size of the detector (Van de Grind, Koenderink and Van Doorn, 1986; Mikami, Newsome and Wurtz, 1986), we hypothesize that this effect may also reflect the distribution of the detectors for that displacement size in the visual field. Figure 3.7f might then indicate that there are more detectors for this spatial displacement size covered by the central 0.5 degree strip of stimulus (figure 3.7e) than the two 0.5 degree strips flanking the central 1.0 degree strip of stimulus even though the former is half of the latter in total area. Thus the data may be a result of a non-uniform distribution of responding motion detectors under the 2x2 degree central visual area covered by our stimulus. If this is true, the central 0.5 degree strip stimulates more detectors tuned to this motion than the two outer 0.5 degree strips. Testing such an explanation requires either further experimentation or a quantitative model that is beyond the scope of the current topic.

A final point of note in figure 3.7 is that the addition of the T_S = 22 ms sample points appears to have improved the accuracy of the reconstruction of the threshold surface with respect to the influence of temporal contrast sensitivity falloff as T_S is decreased. This improvement is shown in the change of contour shape between figures 3.7a and 3.7c. The addition of sample points for N_D = 4 also aids accuracy in the reconstruction of the sampled surfaces because that portion of the threshold surface changes from being flat (figure 3.7c) to having some variation in that region of the graph (figure 3.7f). For example, the LSNR = 1 contour moves into the center region of the contour diagram as the display configuration is changed from that in figure 3.7c to that in figure 3.7f.

3.4 General Discussion

The outcome of the experiments presented here can be related to a wide range of previous reports, but we restrict ourselves to those experiments that are most closely related in terms of stimulus type and intended line of investigation. We are especially interested in relating our results to experiments that seem to indicate that temporal integration in the perception of apparent motion is governed by N_D rather than D_S since our data provide evidence conflicting with that conclusion. We begin with previous work on the extension of the maximum perceivable spatial displacement by Snowden and Braddick (1989a). This quantity is often referred to as D_{max} (Braddick, 1973) in two frame apparent motion

experiments. The analogous (and quantitatively very similar) measure in our experimental paradigm is S_{max}^{100}. Their first experiment (their figure 1) showed that D_{max} increased by a factor of about 1.5 as N_D was increased, reaching the final value after 4-6 displacements. Their stimulus was somewhat different from that used here: a 3.7x3.7 degree stimulus, T_S = 10 ms, ISI = 30 ms, using the method of adjustments to track a 71% correct left-right directional motion perception threshold. The experiment employed $N_L > 1$, but this did not interfere with the measurement of D_{max} because no correlation can be perceived for spatial displacements of $2xD_{max}$ or greater; i.e., no spatio-temporal correlation is possible between non-temporally adjacent image frames under those experimental conditions. They show this in their experiment 3, figure 6, using a stimulus similar to ours but using a spatial signal-to-noise ratio (SSNR) threshold measurement method (changing the ratio of coherently to incoherently displaced dots). They used an ISI of 30 ms with T_S = 10 ms, so comparison with our data would be most appropriate at T_S = 40 ms, a value that we did not directly measure. Our data show a similar amount of improvement and the same asymptotic behavior for $T_S \geq 33$ ms but indicate that the absolute amount of improvement depends on the value of T_S.

Their second experiment was intended to test whether the improvement in D_{max} was governed by a) a velocity limit, b) a limited integration time, c) a spatial integration limit, or d) a spatial displacement count limit. To do this they varied stimulus onset asynchrony[1] and compared the resulting threshold curves to hypothetical curves that would result from each of these temporal integration mechanisms. They conclude that the results look like the hypothetical curves representing either c and d above (their figure 2c). They reached this conclusion in part because they attributed a lowered T_S = 20 ms curve (shifted to lower sensitivity) to a poorly processed first displacement. Our results show that the lowering of the 20 ms curve can be accounted for by temporal contrast sensitivity loss due to their 50 Hz frame rate. Their results are roughly equivalent to the S_{max}^{100} curves in our figure 3.5 (with their parameter values falling between those in our figures 3.5a and 3.5b), but because we used a wider range of stimulus parameters, our conclusions differ.

We must also consider the results of experiments in Snowden and Braddick (1989b). In their previous publication (Snowden and Braddick, 1989a) the authors concluded that N_D was the parameter of interest, and this follow-up work was intended to further analyze that conclusion[2]. Their first two experiments measure several threshold curves for a 7 arc-min (about 0.12 degree) spatial displacement size over various T_S values (20, 50 and 100 ms with no ISI) and N_D values (1-8) using a stimulus 3.7x3.7 degrees in size. If the claim that their 50 and 100 ms curves remove correlation between temporally non-adjacent image frames (i.e. frames 1 and 3, 1 and 4, etc.) is true[3], then the T_S = 50 and T_S = 100 curves in their figure 1 and the T_S = 20 ms curves in our figure 3.3 are equivalent to the corresponding constant T_S curves in our figure 3.6f or 3.6g. Inspection of our figure 3.6f reveals many similarities. Both sets of data show threshold saturation with increasing N_D for a constant T_S, and a similar threshold saturation level for $T_S > 33$ ms. The lower saturation limit for their T_S = 20 ms curve is also reflected in our data, a result that can be accounted for by temporal contrast sensitivity loss.

---

[1] Stimulus onset asynchrony (SOA) is the total time between the onset of each spatial pattern exposure and includes the interstimulus interval, or the amount of time that a blank pattern is shown between spatial pattern exposures. Their second experiment used an ISI in two of the 5 curves, but the ISI did not seem to affect the curve shape.

[2] They discuss their results in terms of cooperative processes mediated by either hetero or homo recruitment (cooperation of detectors with either different (hetero) or similar (homo) span and delay values). In this paper we are not testing any temporal mechanism specifically, so we are more interested in comparing experimental results, but we note that they show very nicely (see their figures 1 and 3) that N_L (dot life time) influences the rate and magnitude of threshold improvement with N_D for small T_S values.

[3] A more general statement is that we only need $S_S \geq D_{max}/2$ (or, in our terms, $S_S \geq S_{max}^{100}$) to prevent visibility of spatial correlation across non-temporally-adjacent frames. From the results in Fredericksen et. al. (in press) we can calculate exactly which $<S_S,T_S>$ combinations can result in spatio-temporal correlations between non-adjacent stimulus frames. This relation also depends on eccentricity of the stimulus in the visual field (op. cit.) and thus the size of the visual stimulus.

Their data show threshold curve variations across T_S that can also be accounted for in our results. Subject RS in their figure 1 shows a loss of sensitivity as T_S is increased from 50 to 100 ms. They use a larger stimulus and a lower dot density, and do not report the overall luminance of the display, so we cannot estimate the influence of threshold saturation (e.g. subject dependent threshold saturation shown for subject EF in our results). However our results reveal the presence of temporal tuning (our figures 3.5 and 3.6) with respect to T_S, and they reveal that the visibility of this tuning depends on the saturation of the threshold. Increasing T_S results in a loss of sensitivity for at least some $<S_S, T_S>$ combinations, and the loss of sensitivity seen for subject RS at T_S = 100 ms in their results can be accounted for by this effect. This comparison may indicate that phenomena visible with our experimental method are also visible using the SSNR technique, although that method runs into problems for very low threshold ratios (see Fredericksen et. al, in press; chapter 2).

We must also consider a third study that uses N_D as a primary parameter and concludes that temporal integration for large and small S_S is mediated by different mechanisms (Snowden and Braddick, 1990). Their intention was to test whether D_{min}, the smallest perceivable spatial displacement analogous to D_{max}, improves with N_D in the same manner as D_{max}. Their first experiment (their figure 1) used a stimulus size of 0.95 x 0.95 degrees, N_L = 1, and T_S = 100 ms. Directional perception error curves for different S_S values were measured for N_D values of 1, 2 and 3. Comparison with our data requires selection of the 21% error position on these curves and plotting that against N_D. They report that there is no significant change in the curves as N_D is increased to 2. Our own data show that, under the same conditions and for some individuals, there is little improvement of S_{min}^{100} (e.g. our figure 3.2d, N_D = 1 and N_D = 2 curves at LSNR = 100) but that if N_D is further increased S_{min}^{100} can be improved.

In their third experiment (their figure 3) they claim to show that stimulus display size does not influence the rate of improvement of D_{min}. They perform an indirect test by measuring D_{max} using the smaller display size from their first experiment to see if the improvement in Dmax is removed by the change in stimulus size. The results show that D_{max} improves as N_D goes from 1 to 2 for the smaller display size. A more direct, and perhaps more justifiable, test would have been to repeat experiment 1 with a larger field size. We assume that this was not done because the field size increase would have required moving the display closer to the subject, thereby reducing the testable range of small S_S values (a restriction also encountered in our own experiments). The dependence of motion sensitivity on stimulus display area has been previously shown (van Doorn and Koenderink, 1984), and our experiment 2 extends this result to stimuli with N_L = 1 and various D_S values. Although our experiment 2 was performed for an S_S value just greater than S_{min}^{100} we would be surprised if this dependence of threshold on stimulus area did not hold for the improvement of S_{min}^{100} (i.e. D_{min}). Note that we must also consider the implications of detector distributions in the visual field as mentioned with respect to experiment 2. We have hypothesized that the results of experiment 2 (figures 3.7e and 3.7f) indicate a distribution of detectors for S_S = 0.0313 degrees that concentrates more detectors in the central portion of the 2x2 degree region of central vision than the outer (upper and lower) regions. If this holds for the smallest perceivable S_S values, that detector distribution can influence the amount of threshold change that results from changing stimulus display area.

Perhaps the most interesting data to compare to our current results is their experiment 4 (their figure 4). The SSNR method is used to measure directional perception thresholds for several S_S values at T_S = 100 ms. In this experiment, however, they use a random selection rule to decide which dots to coherently displace. This means that N_L was not a constant, but rather a random distribution across the coherently displaced dots (each dot could have a different N_L value). The distribution of N_L would depend on the random function used to select the dots for displacement. More importantly, the distribution would also change with the SSNR value of the stimulus. The reasoning behind this choice is not explained, but we assume that the authors were trying to provide a stimulus that would give wide coverage of the perceivable S_S values. Such a choice means that the stimulus did not isolate detector

populations by span, and that interpretation of the threshold within the experimental frame work presented here is difficult[4]. However, although they use the SSNR threshold measurement method (we have previously argued that the SSNR and LSNR methods may determine perceptual thresholds differently [Fredericksen et. al., in press; chapter 2]), comparison of the results of their experiments with those presented here is still valuable. The data presented in their figure 4 would roughly correspond to the $N_D = 1$ and $N_D = 2$ curves in figures 3.2d and 3.3d of our results. This comparison reveals similar behavior in that thresholds for small S_S values show less improvement than thresholds for large S_S values as N_D goes from 1 to 2. The range of S_S values covered in their figure 4 is larger than presented here because their stimulus was larger in size.

In their final experiment (their figure 5) the authors repeat a procedure introduced by Nakayama and Silverman (1984) in which the first and last frame exposure times of a multiple displacement stimulus are held constant while the middle frame exposure duration is varied. The original experiment was designed to test for a minimum displacement processing time (op cit.) and used a random dot shearing motion between spatially interleaved panels in the stimulus. In the reproduced experiment the authors use the SSNR method (apparently with the random dot selection rule) to measure a threshold using $N_D = 2$ (three image frames), and using $T_S = 100$ ms for the first and third exposure durations. Although we must keep the differences in experimental design in mind, the range of stimulus parameters tested in their experiment 5 may be roughly comparable to constant S_S slices through the surface in our figure 3.4b. The comparison shows remarkably similar behavior (specifically, we refer to their figure 5a) between their results and ours. The large and small S_S values in their experiment would fall on the upper and lower bounds of S_S visibility (our figure 3.4) and would thus not reflect the sensitivity optimum and T_S tuning evidenced in the central S_S value that they tested.

[4] The distribution of information across S_S values in their stimulus depends on SSNR, the basic S_S size and the random selection rule. The motion detection threshold can be influenced by each of these factors.

A final comparison can be made, if we again keep in mind the differences in stimuli, with the work of Nakayama and Silverman (1984) (described above). In order to make a quantitative comparison, we would need to compensate their measured D_{max} curves for the value of N_D used to measure that curve (because, as previously mentioned, no spatio-temporal correlation can take place for total displacements larger than D_{max}). Alternatively the *shape* of the curve can be directly compared. Following this guideline, their figure 2 data presenting the D_{max} curve for $N_D = 2$ is (as above) comparable to the S^{100}_{max} contour line in our figure 3.4b. Again, the shape of their curve and ours is similar. Interpretation of their other experiments requires an increasing number of assumptions of equivalence and are thus less valuable, but we would like to point out that in another experiment (their figure 3) they hold $N_D = 2$ and increase the size of the stimulus. The shape of these D_{max} curves reflects both temporal integration (because $N_D = 2$) and how S^{100}_{max} changes with eccentricity of the stimulus in the visual field (Fredericksen et. al, in press; chapter 2). If temporal integration proceeds in a similar manner at points in the visual field outside of central vision (a hypothesis that is not tested here, but seems reasonable), then their figure 5 reflects a combination of the S^{100}_{max} curves in figure 4b of Fredericksen et. al. (in press; chapter 2), the consequences of contrast sensitivity loss at high frame rates, and dependence of threshold on total stimulus duration.

3.5 Physiological Considerations

In searching for a physical model of the temporal integration behavior of the human motion detecting machinery, we should look to the neurophysiology. It is reasonable to assume that perceptual threshold is inversely proportional to sensitivity and that motion detector activity is a monotonic function of the sensitivity of that detector to a given stimulus. If we consider the detector as a spatio-temporal filter, the output can be treated as a measure of the energy collected by the filter. If this energy (activity) is then transmitted to a later stage[5] for temporal integration, we

[5] Temporal integration could also take place in the correlator (simple leaky integration is a linear function), but for simplicity we consider it to be a separate stage.

must consider the possible response properties of that later stage. Leaky integrator components have previously been used in a model of alpha-stripe[6] generation in the visual perception of motion in random dot patterns (Adler, Bock, and Grüsser, 1981). The leaky integration, thresholding, and response saturation behavior of neural cells is well established, and we suggest that these are reasonable components for use in modeling our temporal summation mechanism. The results shown above would seem to support such a model. Specifically it appears that in central vision temporal integration depends on stimulus duration regardless of spatial displacement size but is also influenced by temporal contrast sensitivity loss at high frame rates and a T_S tuning behavior that is inversely proportional to S_S. The output of a leaky integrator depends on the shape and magnitude of the input to the integrator over time. Contrast sensitivity and T_S tuning both influence the activity of the motion detector, thus influencing the subsequent leaky integration of the detector's output. Direct tests of this model will be reported elsewhere (Fredericksen, Verstraten and Van de Grind, in preparation).

3.6 Conclusions

Contrary to previous results our data show that both S^{100}_{min} and S^{100}_{max}, in fact the thresholds of all perceivable S_S values in these experiments, can be improved through temporal integration of the motion stimulus. Sensitivity at all $<S_S,T_S>$ combinations measured here can be improved by increasing N_D. The amount of improvement is dependent on $<S_S,T_S>$ as well as subject-specific factors. Sensitivity can be predicted by a constant N_D in some regions of the <S,T> plane, but those are regions where the response saturation time limit is easily reached within a small number of T_S increments. This explains previous conclusions that N_D is the primary stimulus parameter determining the saturation of temporal integration in directional motion perception. Our results indicate that D_S is a better predictor of threshold improvement in apparent motion than is N_D over the range of $<S_S,T_S>$ values tested here, but spatio-temporal tuning of the detectors, contrast sensitivity loss due to frame rate, spatial summation and subject dependent factors (e.g. saturation effects) all affect that threshold. The similarities of the hypothetical surfaces in figures 3.6a and 3.7b to the measured threshold data provide evidence that a model based on total stimulus duration combined with the influence of temporal contrast sensitivity captures some major temporal integration properties of the motion detecting system.

Results on the dependence of S^{100} on eccentricity of the stimulus in the visual field (Fredericksen et al., in press; chapter 2) indicate that the variation of S^{100}_{min} and S^{100}_{max} across the visual field can be accounted for by the properties of known visual anatomy in a uniform manner for both S^{100}_{min} and S^{100}_{max}. The uniformity of the behavior of the human motion detection system with respect to both temporal summation and the range of perceivable S_S values (as indicated by our current and previous experimental results) provides compelling evidence that the perceivable range of $<S_S,T_S>$ values are implemented in a single, consistent manner yielding a set of detectors with a uniform architecture but using a set of components with varying spatio-temporal tuning characteristics. These varying spatio-temporal tuning characteristics are probably a result of heterogeneous temporal receptive field properties of the components of the detector.

Acknowledgments

R. E. Fredericksen was supported by a UNC Cognitive Science Program Pre-doctoral Fellowship, and an International Brain Research Organization Research (IBRO) Fellowship. F. A. J. Verstraten was supported by the Biophysics Foundation of the Netherlands Organization for Scientific Research (NWO). We thank Marieta Braks for serving as an observer. Special thanks go to Peter Schiphorst and Alexandra Bellekom for designing and building the custom stimulus generation hardware, and to Rob van Weerden for taking care of the 1001 peripheral details required for use of the equipment during experimentation.

[6] If a subject views random dot pattern apparent motion with a fixed point of gaze, then for some frame rates a stationary striation of the random dot pattern is seen perpendicular to the direction of motion. These striations are called alpha-stripes, and represent an internally generated modulation of the percept.

Bibliography

Adler, B., Bock, O., and Grüsser, O.-J. (1981) Alpha-stripes: A spatial periodicity appearing in stroboscopically illuminated moving random dot patterns. *Vision Research*, **21**, 913-924.

Braddick, O. (1973) A short-range process in apparent motion. *Vision Research*, **14**, 519-527.

Burr, D.C., (1981) Temporal summation of moving images by the human visual system. *Proceedings of the Royal Society of London*, **B 211**, 321-339.

Fredericksen, R. E., Verstraten, F.A.J. and Van de Grind, W.A. (submitted), An analysis of the mechanism underlying temporal integration of motion information. Submitted.

Fredericksen, R. E., Verstraten, F.A.J. and Van de Grind, W.A., (in press) Spatio-temporal characteristics of human motion perception, to appear in *Vision Research*.

Kelly, D.H. and Burbeck, C.A. (1984) Critical problems in spatial vision, in CRC Critical Reviews in Biomedical Engineering, **10:2**, 125-177.

McKee, S. P. and Welch, L. (1985) Sequential recruitment in the discrimination of velocity. *JOSA A*, **2**, 243-251.

Mikami, A., Newsome, W.T., and Wurtz, R. H. (1986). Motion selectivity in macaque visual cortex. I. Spatiotemporal range of directional interactions in MT and V1, *Journal of Neurophysiology*, **55:6**, pp. 1308-1327.

Mikami, A., Newsome, W.T., and Wurtz, R. H. (1986). Motion selectivity in macaque visual cortex. II. Mechanisms of direction and speed selectivity in extrastriate area MT, *Journal of Neurophysiology*, **55:6**, pp. 1328-1339.

Mikami, A., Newsome, W.T., and Wurtz, R. H. (1986). Motion selectivity in macaque visual cortex. III. Psychophysics and physiology of apparent motion, *Journal of Neurophysiology*, **55:6**, pp. 1340-1351.

Nakayama, K. and Silverman, G. H. (1984) Temporal and spatial characteristics of the upper displacement limit for motion in random dots. *Vision Research*, **24**, 293-299.

Reichardt, W. (1961) Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. In Sensory Communication, Rosenblith, ed. (Wiley, New York, 1961), pp. 303-317.

Scobey, R. P. and Kan, P. L. E. (1991) A horizontal stripe of displacement sensitivity in the human visual field. *Vision Research*, **31**, 99-109.

Snowden, R. J. and Braddick, O. J. (1989a) Extension of displacement limits in multiple-exposure sequences of apparent motion. *Vision Research*, **29**, 1777-1787.

Snowden, R. J. and Braddick, O. J. (1989b) The combination of motion signals over time. *Vision Research*, **29**, 1621-1630.

Snowden, R. J. and Braddick, O. J. (1990) Differences in the processing of short-range apparent motion at small and large displacements. *Vision Research*, **30**, 1211-1222.

Snowden, R. J. and Braddick, O. J. (1991) The temporal integration and resolution of velocity signals. *Vision Research*, **5**, 907-914.

Van Doorn, A.J., and Koenderink, J. J. (1982a) Spatial properties of the visual delectability of moving spatial white noise. *Experimental Brain Research*, **45**, 189-198.

Van Doorn, A.J., and Koenderink, J. J. (1982b) Temporal properties of the visual delectability of moving spatial white noise. *Experimental Brain Research*, **45**, 179-188.

Van Doorn, A.J., and Koenderink, J. J. (1984) Spatiotemporal integration in the detection of coherent motion. *Vision Research*, **24**, 47-53.

Van de Grind, W. A., Koenderink, J. J. and Van Doorn, A.J., (1986) The distribution of human motion detector properties in the monocular visual field. *Vision Research*, **26**, 797-810.

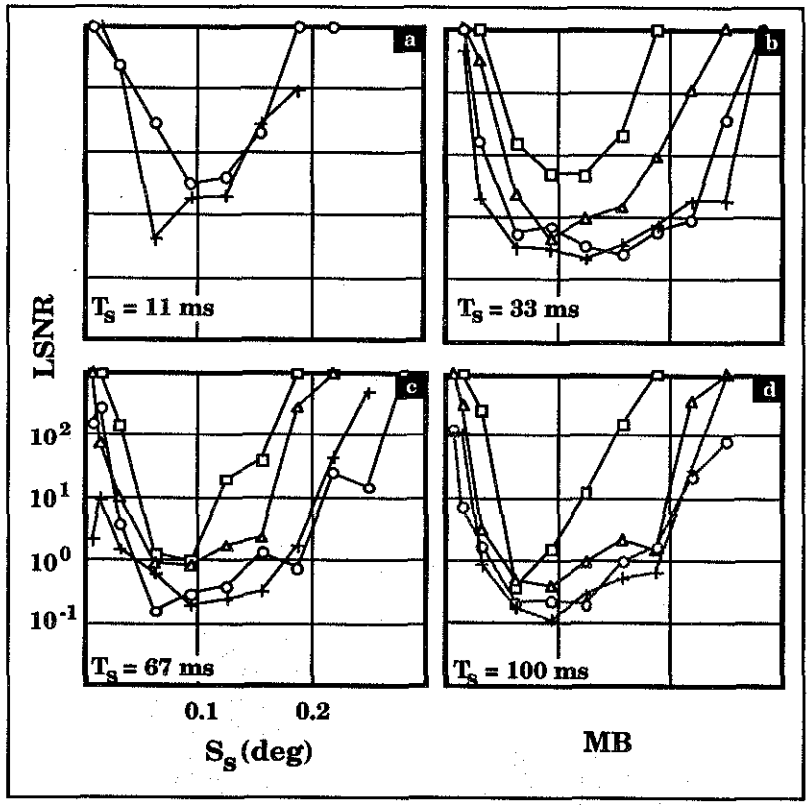
Van de Grind, W. A., Koenderink, J. J., Van Doorn, A.J., Milders, M.V., and Voerman, H. (in press) Inhomogeneity and anisotropies for motion detection in the monocular visual field of human observers. To appear in *Vision Research*.

Van de Grind, W.A., Koenderink, J.J., and Van Doorn, A.J. (1987) Influence of contrast on foveal and peripheral detection of coherent motion in moving random-dot patterns. *JOSA A*, **4**, 1643-1652.

Van de Grind, W. A., and Grüsser, O.-J. (1973) Temporal transfer properties of the afferent visual system, in Handbook of Sensory Physiology, Volume VII/3A, R. Jung, editor.

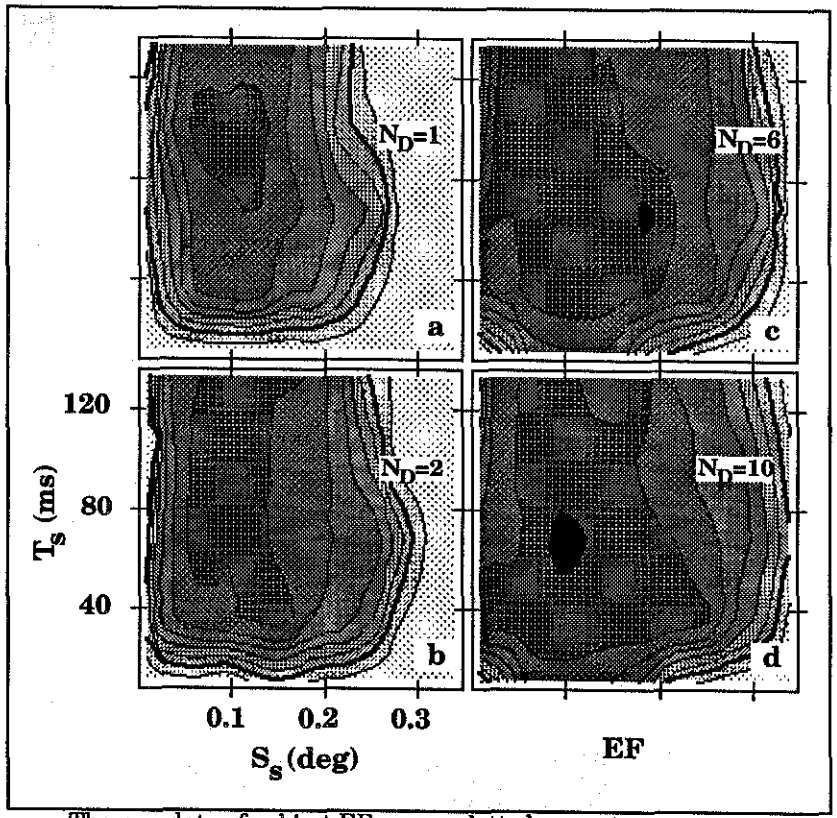
Appendix

Raw data for other subjects

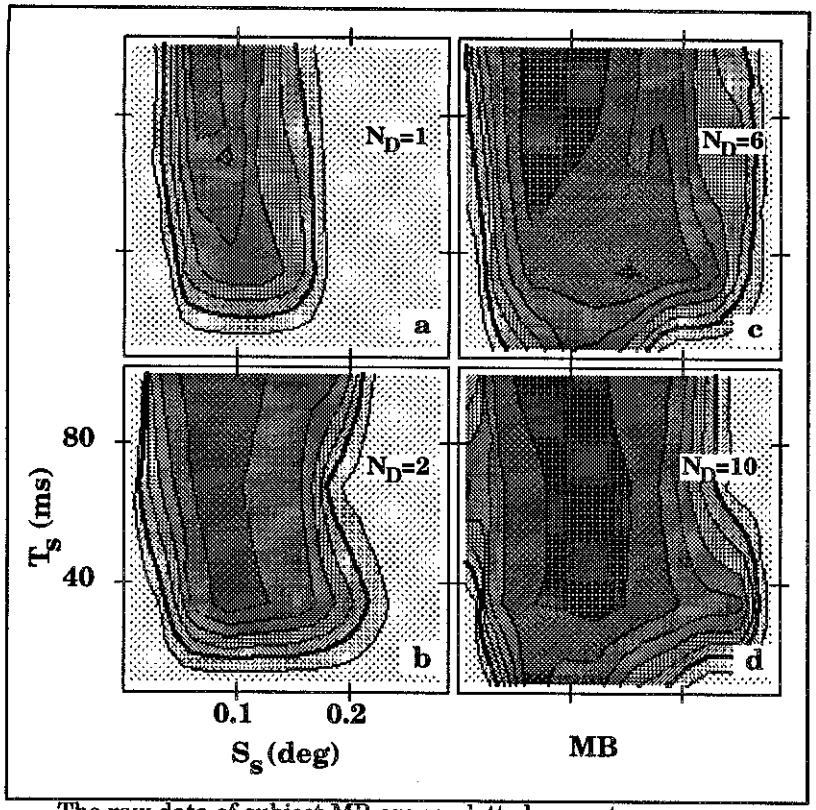


Raw data for subject MB are plotted as in figures 2 and 3. Each graph displays a set of LSNR threshold curves as a function of S_S. The parameter of the curves within each graph is N_D. Unique symbols are assigned to represent a constant displacement count: □ for N_D=1, Δ for N_D=2, o for N_D=6, and + for N_D=10. The parameter across the graphs is T_S and each graph is labeled accordingly.

Constant N_D surfaces for other subjects

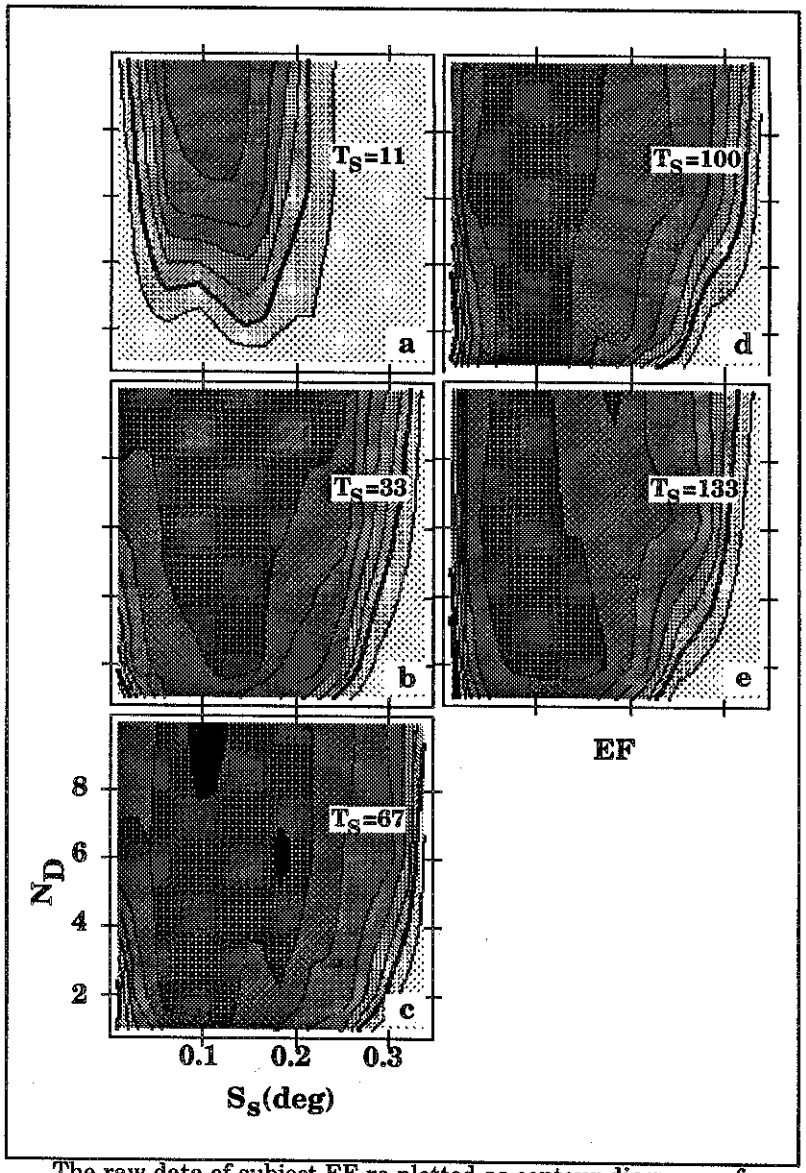


The raw data of subject EF are re-plotted as contour diagrams of the threshold surfaces obtained by fixing the value of N_D. Contours are shown every half log unit with the S^{100} contour drawn with a wider pen. Dark areas indicate low LSNR values (high sensitivity) while light areas indicate high LSNR values (low sensitivity).



The raw data of subject MB are re-plotted as contour diagrams of the threshold surfaces obtained by fixing the value of N_D. Contours are shown every half log unit with the S^{100} contour drawn with a wider pen. Dark areas indicate low LSNR values (high sensitivity) while light areas indicate high LSNR values (low sensitivity).

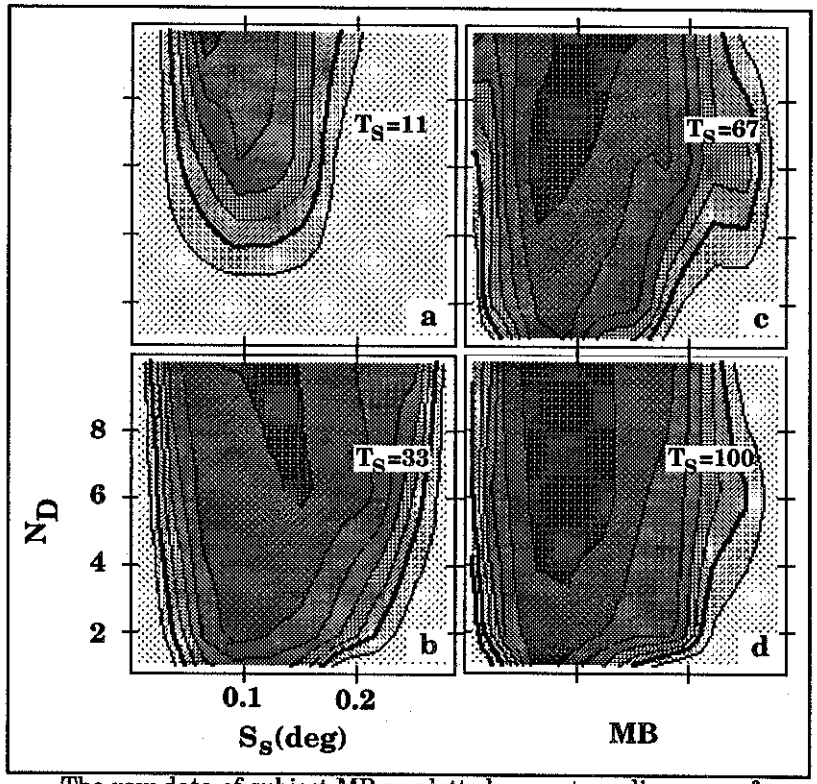
Constant T_s surfaces for other subjects



The raw data of subject EF re-plotted as contour diagrams of the threshold surfaces obtained by fixing the value of T_s.

Contours are shown every half log unit with the S^{100}

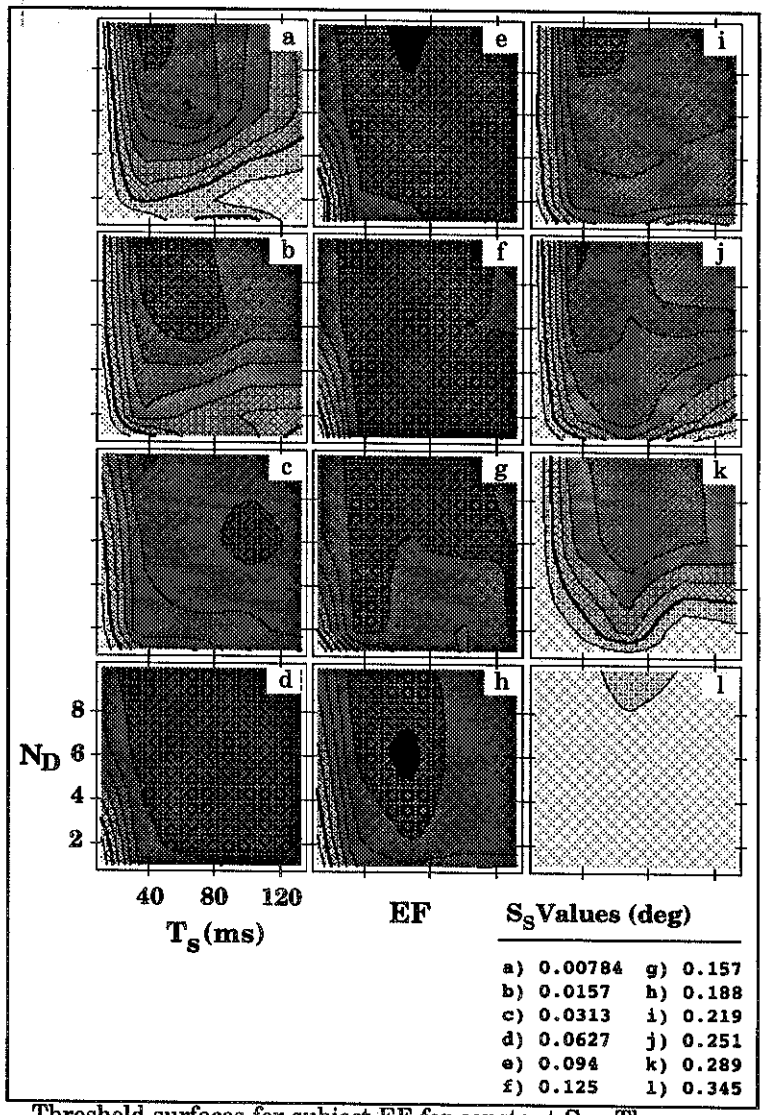
contour drawn with a wider pen. Dark areas indicate low LSNR values (high sensitivity) while light areas indicate high LSNR values (low sensitivity).



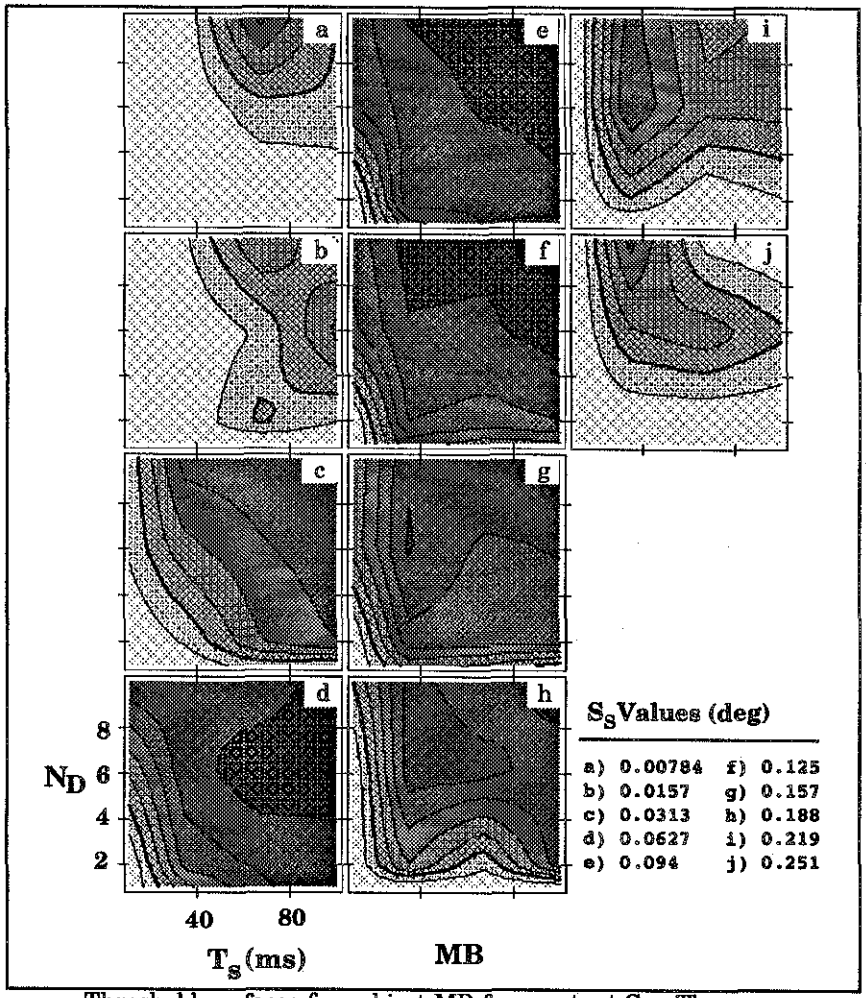
The raw data of subject MB re-plotted as contour diagrams of the threshold surfaces obtained by fixing the value of T_S.

Contours are shown every half log unit with the S^{100} contour drawn with a wider pen. Dark areas indicate low LSNR values (high sensitivity) while light areas indicate high LSNR values (low sensitivity).

Constant S_S surfaces for other subjects



Threshold surfaces for subject EF for constant S_S. These LSNR threshold surfaces over the $<T_S,N_D>$ plane are plotted in the same manner as in figures 3.4 and 3.5. The legend included in the lower right hand corner of the figure indicates the fixed S_S value in each graph.



Threshold surfaces for subject MB for constant S_S. These LSNR threshold surfaces over the $<T_S,N_D>$ plane are plotted in the same manner as in figures 3.4 and 3.5. The legend included in the lower right hand corner of the figure indicates the fixed S_S value in each graph.

Chapter 4*: An analysis of the temporal integration mechanism

R. E. Fredericksen, F. A. J. Verstraten and W. A. van de Grind

Abstract

We present a model for the temporal integration of random dot apparent motion. The model is constructed from psychophysical threshold measurements presented here as well as other psychophysical and neurophysiological data. The mechanism is based on the leaky integration of pulsatile motion detector responses to apparent motion stimuli. The pulses represent the motion detector's responses to the discrete spatial displacements in the apparent motion stimulus. Temporal contrast sensitivity determines the shape of constant-stimulus-duration threshold curves for image frame rates greater than about 7.5 Hz. The shape of the threshold curve for frame rates less than 7.5 Hz is determined by the leaky integrator time constant and the shape of the pulses emitted by the motion detector. The leaky integrator model exhibits threshold saturation behavior seen in psychophysical data as well as the dependence of threshold saturation time on the frame rate of the apparent motion stimulus. A low frame rate results in a longer saturation time because the leaky integrator discharges more between pulses. When the motion detector output pulses are far enough apart, there is effectively no temporal integration and thus no threshold improvement over time. This behavior corresponds to the lack of temporal integration in psychophysical threshold measurements under the same psychophysical conditions. Finally, the behavior of the psychophysical threshold curves across spatial displacement sizes indicates the influence of a response threshold mechanism that is an integral part of the temporal integration model.

* The material in this chapter will be submitted to the journal *Vision Research* in almost identical form under the title "An analysis of the temporal integration mechanism in human motion perception" by the above authors.

4.1 Introduction

Objects in our physical world interact with light energy that impinges on them by re-emitting and/or reflecting spatio-temporal patterns of light. Many animal visual systems have evolved so as to be able to intercept these spatio-temporal patterns of energy and compute from them measures of the objects that are useful to our survival. In recent years we have learned to manufacture spatio-temporal patterns of light energy, using video displays, that mimic the information that comes from the interaction of a "real" object with light. These video displays have become increasingly widespread in the study of visual perception and are powerful tools, in part due to their great flexibility. The spatio-temporal pattern of energy that is produced when an object moves through the world, for example a tree falling in the forest or a bird flying through the air, is influenced by the relative position and motion of the object and the position of the light source(s). Our own visual systems have acquired the ability to intercept the resulting spatio-temporal patterns of light radiation, transduce them into neuronal signals, and then successfully compute from those signals an estimate of the object's speed and direction of motion relative to ourselves and the light source (if the relationship is not too complicated). The motion information that results from an object's interaction with light sources as it moves relative to them changes continuously over time, and the specific portion of the spatio-temporal energy pattern that an organism can intercept depends on the relative positions of the organism and the object.

However, in the study of motion perception a video display creates additional problems. Motion information created using computer driven video displays, called apparent motion, is a display of a sequence of essentially static images, snapshots of the original scene, that we perceive as motion, even though there is no "real" object moving in our field of view. The spatio-temporal pattern of energy that our eyes intercept mimics a "real" object that is moving through the world in a sequence of *instantaneous spatial translations* rather than in a smooth arc. The spatio-temporally discrete nature of apparent motion influences our perception of

the motion, but we perceive it as motion nonetheless. If we wish to "convince" the visual system of a person viewing such a sequence of images that the apparent motion is real, then we must keep certain parameters of the stimulus within known, well defined ranges. However the discrete nature of an apparent stimulus can also provide information that will allow us to discover something about how our visual systems compute visual motion information. It is the temporal aspect of apparent motion stimuli that is perhaps the most anomalous, and the subject of a great deal of investigation. A very short list of some of the temporal properties of apparent motion perception that have been investigated includes the required sampling frequency for the perception of smooth motion (Morgan, 1979; Burr, Ross and Morrone, 1986), the effects of exposure duration and interstimulus interval in the perception of motion in two-frame random dot pattern sequences (Baker and Braddick, 1985; Van Doorn, Koenderink and Van de Grind, 1985), and perception threshold improvement with motion stimulus duration (Burr, 1981; Van Doorn and Koenderink, 1984; Nakayama and Silverman, 1984; McKee and Welch, 1985; Snowden and Braddick, 1989a, 1989b, 1991). It is this last property that motivates the work presented here.

The visibility of a motion stimulus improves with the amount of time that the motion information is available to the visual system. The improvement of the visibility begins at the first moment that the motion begins and eventually reaches some maximum visibility. The term *visibility* can be defined in terms of a psychophysical measurement of the threshold for, or sensitivity to, the stimulus. The psychophysical threshold measurement is defined to be inversely proportional to the psychophysical sensitivity measurement, and the sensitivity measurement corresponds to the strength of the signal produced by the biological machinery that computes the motion information. In terms of a psychophysical measurement, the signal is the aggregate output of the population of motion detectors responding to the motion stimulus. We will use the term *response saturation* to refer to the maximum signal strength that is reached by the motion detector population over time. Response saturation refers to the point in time at which psychophysical sensitivity (and thus threshold) no longer improves. The terms response saturation, sensitivity saturation, and threshold saturation will be used interchangeably throughout the remainder of the paper.

In previous investigations of temporal improvement of apparent motion thresholds both stimulus duration[1] (D_S) and spatial pattern displacement count (N_D, the number of times that a spatial pattern is rigidly translated on the video display) have been used as a independent variables (Nakayama and Silverman, 1984; McKee and Welch, 1985; Snowden and Braddick, 1989a, 1989b, 1991). Some experimental results seemed to provide evidence that N_D, rather than D_S, is the variable that controls the saturation of sensitivity to an apparent motion stimulus (Snowden and Braddick, 1989a). In a separate publication we present evidence that D_S in combination with motion detector populational tuning characteristics provides a more precise measure of temporal improvement of directional apparent motion thresholds in humans than does N_D although under certain stimulus conditions N_D can provide an approximate measure of temporal improvement of motion sensitivity (Fredericksen, Verstraten, and Van de Grind, submitted; chapter 3). In our previous paper we propose a mechanism for the improvement of directional motion thresholds over time[2] that is based on known neurophysiological properties. The data presented there are consistent with that model, which consists of leaky integration of motion information over time with response saturation and response thresholding. We stress here that the term *response threshold* is used in this context to refer to an absolute detector response magnitude which must be reached for reliable detection of the stimulus to occur. The terms response magnitude, response threshold and response saturation refer to the aggregated signals of the population of motion detectors, while the terms sensitivity and threshold refer to psychophysical measures of the aggregate signal which reflect the populational properties.

---

[1] We use the term **stimulus duration** in apparent motion to mean the total time that spatio-temporal motion information is presented to the viewer. We do not include the presentation of uniform luminance fields before the first or after the last frames of the image sequence, but must include any interstimulus intervals (presentation of uniform luminance patterns) between image frames.

[2] The process of improving the perception of motion over time has been labeled both temporal integration and temporal summation. We will use the term temporal integration throughout the paper because of the nature of the proposed mechanism.

The experiments and simulations presented here were designed to further explore the hypothesized leaky integration mechanism and to extend the model of temporal integration presented in our previous paper. The results of the previous experiments indicate that D_S is a more precise indicator of threshold improvement, so we first present threshold results for the detection of directional motion taken along that dimension. The second set of psychophysical experiments was designed to elucidate the influence of temporal contrast sensitivity in the detection of directional motion. The results of these psychophysical experiments are used as a guide in creating a model for temporal integration of motion information based on the leaky integration of motion detector output responses over time. The results of a simulation of the model are presented, and the correspondences between the model and the current and previous threshold results are discussed. The leaky integrator model captures the threshold saturation behavior seen in the psychophysical data, as well as the dependencies of threshold saturation time constants on image frame duration. The experimental results indicate the influence of a response threshold determined by the maximum average contrast of the stimulus.

We begin by giving a summary of the stimulus and threshold measurement paradigm used in our current and previous work. The stimulus is composed of a sequence of spatially displaced random dot patterns displayed on a video display screen. The stimulus was designed with a functional model of human motion perception as a guide (figure 4.1). The model employs a front end array of correlational devices (bi-local detectors in figure 4.1) the outputs of which are used in higher level computational stages. In the current experiments we are interested in the higher level computational function of temporal integration. The random dot apparent motion stimulus was designed to isolate specific populations of the spatio-temporally tuned motion detectors. Isolation of motion detector populations is also important here because it can reveal any differences in the temporal integration mechanism across the detector population.

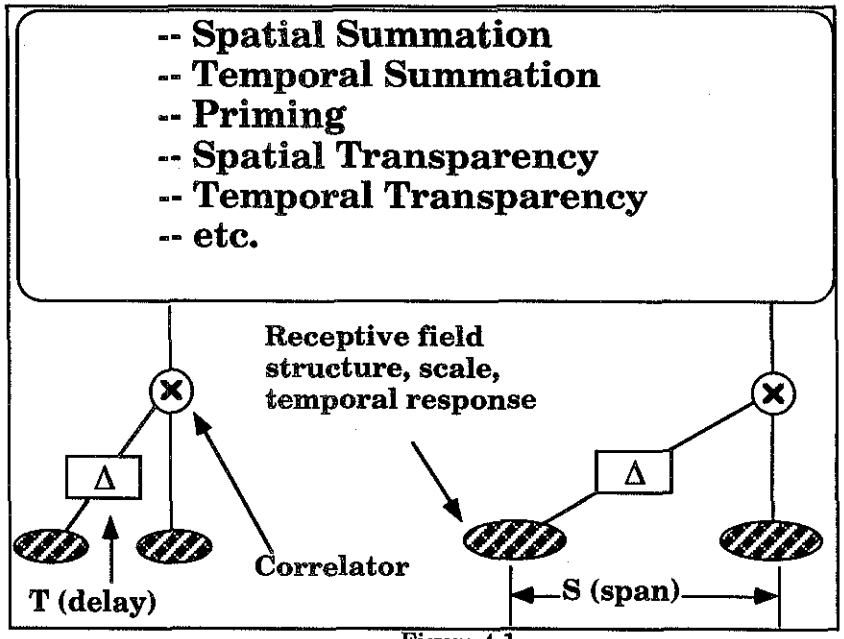


Figure 4.1
The functional performance of the human visual motion perception system is modeled as an array of bi-local detectors whose outputs are fed to later computational stages. The detector array can be characterized by distributions of detector span (distance between receptive field centers) and delay (differential delay of the receptive field components) at each position in the visual field. The receptive field components send their output to the correlation component of the detector and can be characterized by their spatial structure, scale and temporal response.

To simplify discussion of the stimulus we define symbols for some stimulus parameters. The frame duration (spatial pattern exposure time) and spatial pattern displacement size in the stimulus are referred to as T_S and S_S, respectively. T_S in the inverse of stimulus frame rate, and we will use the two terms interchangeably. The stimulus duration in our apparent motion stimuli is defined as

$$D_S(T_S, N_D) = T_S \cdot (N_D + 1) \tag{4.1}$$

because we do not use an interstimulus interval (presentation of a uniform spatial luminance pattern between the spatially displaced pattern images).

Dot life time N_L, the number of coherent spatial displacements that a dot in the spatial pattern undergoes before it is randomly displaced, controls the spatio-temporal displacement content of the stimulus. We used $N_L = 1$ in all directional motion detection experiments because it limits the coherent motion information contained in the stimulus to a single $<S_S,T_S>$ combination[3] (for a detailed explanation, see Fredericksen, Verstraten and Van de Grind, in press; chapter 2). Finally, we present threshold results both from directional motion detection experiments that use coherent motion in the stimulus and from non-directional motion detection experiments that use incoherent motion in the stimulus. The directional detection experiments used either a luminance signal-to-noise ratio (LSNR; see Methods) or average stimulus contrast as[4] the threshold variable. The incoherent motion detection experiments used only contrast as the threshold variable. In order to simplify references to these experiments throughout the rest of the paper, we will use the unambiguous terms *directional LSNR threshold*, *directional contrast threshold*, and *detection threshold* to refer to the three kinds of experimental results.

4.2 Methods

4.2.1 Stimulus Generation

The motion stimuli were generated using custom image generation hardware driven by a Macintosh IIfx computer. The monitor had a P4 phosphor and a base display rate of 90 Hz. All motion frame exposure durations were integer multiples of the base frame exposure duration. The interstimulus interval was negligible in all experiments. The display screen was 14 centimeters and 256 pixels square, and pixels were 0.55 mm in diameter. All experiments were performed at a distance of 4.0 meters, so each pixel subtended 0.47 minutes of arc. The display region was 2.0 degrees square in all experiments. A black fixation dot (approximately 2 minutes of arc in width) was affixed to the center of the stimulus display area. Thresholds were determined using either a luminance signal-to-noise ratio (LSNR) method or a contrast method. LSNR threshold measurements were produced by modulating the luminance of the target random dot pattern through addition of a spatially and temporally uncorrelated noise random dot pattern. Some reasons for use of such a paradigm have been enumerated in Fredericksen et al. (in press; chapter 2) and the method has been described in detail elsewhere (van Doorn and Koenderink, 1982a, 1982b). The contrast threshold measurements were produced through manipulation of the average stimulus contrast. Experiments using coherent motion consisted of moving random dot patterns with pixel life times of two frames (one spatio-temporal displacement). The motion noise detection experiment used dot life times of 1 frame and therefore contained no coherent motion, only motion noise. The average luminance of the signal-plus-noise random dot patterns was set to 50 cd-m^{-2} in all experiments. The signal-plus-noise random dot pattern in the LSNR directional motion detection experiment had an average contrast level of 70%, a value well above the motion threshold saturation value of 35% (van de Grind, Koenderink and van Doorn, 1987).

4.2.2 Threshold Measurement

Directional motion detection thresholds were measured using two-alternative-forced-choice horizontal (left-right) motion discrimination so as to distinguish a direction discrimination threshold from a simple motion, contrast change, or flicker detection threshold. Left-right motion was used to discount anisotropies in motion direction discrimination ability (Scobey and Kan, 1991; van de Grind, Koenderink, van Doorn, Milders and Voerman, in press).

Conversely, motion noise *detection* thresholds were measured using two-alternative-forced-choice pattern detection. The motion noise stimulus was presented on either the left or right half of the display, with an equivalent mean luminance (flat luminance profile) displayed on the other half of the

---

[3] When $N_L > 1$ there are spatial correlations of random dots across image frames that are not adjacent in the sequence, thereby providing coherent motion information at integer multiples of the base $<S_S,T_S>$ value. For example, if $N_L = 3$, then there is coherent motion information at $<S_S,T_S>$, $2 \times <S_S,T_S>$, and $3 \times <S_S,T_S>$.

[4] The stimulus is made up of the superposition of a target and a noise pattern. Each pattern has an average (root-mean-square, or RMS) contrast. The signal and noise patterns are spatially and temporally independent, so the average stimulus contrast is $\sqrt{C^2_{signal} + C^2_{noise}}$ where C_{signal} and C_{noise} are the RMS contrasts of the signal and noise patterns, respectively.

display. The task was to choose the side of the display on which the motion noise stimulus was presented.

In both types of experiments the motion stimulus presentation consisted of a number of frames of motion information superimposed on a mean luminance background. The thresholds were determined using a staircase procedure that tracked the 79% correct level; three consecutive correct directional discriminations/position detections resulted in a lowering of the threshold variable value (LSNR or contrast), while any other sequence raised the stimulus variable value. There were 10 turning points in all experiments with the threshold calculated as the average of the final six turning points. A step size of 2 dB was used for the first 4 turning points and a step size of 1 dB was used for the final six. Turning point sequences were inspected, and any sequence that did not reach a stable fixed point due to attentional variation was rejected and re-measured. In order to decrease the required experimental time, the subjects were asked to manipulate the threshold variable manually (before beginning each staircase procedure) until it was just above subjective threshold. If a subject mis-estimated the starting value, this fact was apparent in the turning point sequence of the staircase procedure as a continuing upward trend in the turning points. Such sequences were rejected and the point re-measured. Subjects were allowed to rest after any or all of the individual threshold measurements. The experiments were performed in a portion of a room screened from stray light with ambient lighting provided by the display screen. All experiments were monocular, and the subjects used a chin and head rest. Subjects were instructed to fixate on the black dot on the screen and to determine either the direction of motion or the side of the display on which the stimulus was presented. The subject's answer was recorded via a key press on the computer keyboard, and the threshold variable was automatically changed by the computer. All subjects had normal or corrected to normal vision.

4.3 Directional LSNR Thresholds

4.3.1 Experiment Description

In previous experiments we investigated temporal integration by varying the N_D, T_S and S_S parameters separately (with N_D in mind as the primary parameter) and measuring a directional LSNR threshold for a range of values for each parameter (Fredericksen et. al, submitted; chapter 3). When the resulting threshold data are plotted as a threshold surface on the $<T_S,N_D>$ plane (i.e. for a fixed S_S value), iso-LSNR curves on the threshold surface appear to follow the form that would be predicted by constant D_S lines on the $<T_S,N_D>$ plane. Figure 4.2a shows a contour diagram of a $<T_S,N_D>$ (fixed S_S) threshold surface reproduced from our previous data. The sample spacing in the previous experiments was chosen to be a grid over the plane (shown as small black dots in figure 4.2a) so the sample points were not specifically chosen to be on constant-D_S contours. The first experiment presented here was designed to measure directional LSNR thresholds along the more appropriate dimension of constant-D_S, and to extend the range of sample points.

The positions of the sample points in the previous and current experiments are shown in figure 4.2c. The small black dots are the positions shown on the threshold surface in figure 4.2a, while the larger, lighter gray dots (connected by lines) represent the sample point positions used in the current experiment. Directional LSNR thresholds were measured along these lines representing a constant value of D_S. The combinations of N_D and T_S used were restricted by the discrete nature of the display because D_S and T_S must both be an integer multiple of the basic frame duration of the monitor. The combinations of N_D and T_S for each constant D_S threshold curve were chosen following those restrictions while trying to ensure good sampling of the curve. The directional LSNR threshold curves were taken for various values of S_S selected to cover the range of visible S_S for each subject. Because we were interested in constant D_S rather than N_D or T_S, we were able to measure a wider range of T_S and N_D values than in the previous experiment, allowing a clearer display of the T_S tuning behavior of the population of motion detectors responding to the apparent motion stimulus.

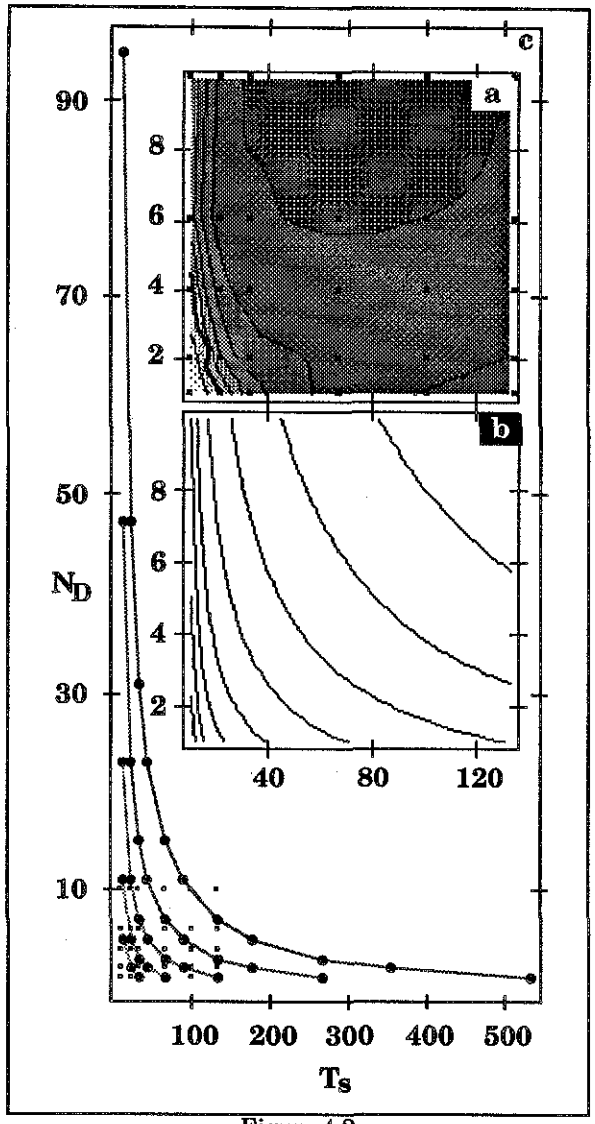


Figure 4.2

The sampling paradigm in the previous, N_D oriented, experiments was a grid over $<T_S,N_D>$ plane. (a) threshold surface taken using the sample-grid with sample points indicated by the small black dots. (b) model threshold surface as defined in the text. (c) a comparison of the old, grid-style, $<T_S,N_D>$ plane sampling method and the current, constant-D_S oriented sampling method. The small black dots are the same as in (1) while the larger, lighter gray dots connected by lines are the sample points used in the current experiments presented here.

4.3.2 Results and Discussion

The results for subjects EF and FV are shown in figures 4.3 and 4.4. Each graph presents threshold curves measured at D_S values of 66 ms, 133 ms, 266 ms, 533 ms, and 1066 ms. All curves in a single graph were measured at a single S_S value (as labeled in the graph). The vertical axis is LSNR threshold and the horizontal axis is T_S.

The data reflect a number of behaviors seen in the previous N_D-oriented experiments. Both subjects show a U-shaped threshold curve for S_S values at or near the perceivable limits of spatial displacement sizes (figures 4.3a, and 4.3e for subject EF and figures 4.4a and 4.4d for subject FV). Subject FV shows less of the U shape threshold behavior, but it is also known that the range of visible S_S values for FV is different than for EF (Fredericksen, et. al., in press; chapter 2). The range of visible S_S values for FV and EF overlap. The S_S values presented in the graphs are absolute spatial displacement sizes. Comparison of the two sets of data require normalization of the absolute S_S values into relative positions in each subject's perceivable range. If we consider the S_S values after scaling them to values between 0 and 1, where 0 and 1 are the minimum and maximum perceivable displacements for the subject, then the S_S value in figure 4.4d (S_S = 0.188) for subject FV is between the S_S values in figures 4.3d (S_S = 0.188) and 4.3e (S_S = 0.313) for subject EF. Likewise the S_S value in figure 4.4d for FV is smaller than the S_S value in figure 4.3e for subject EF.

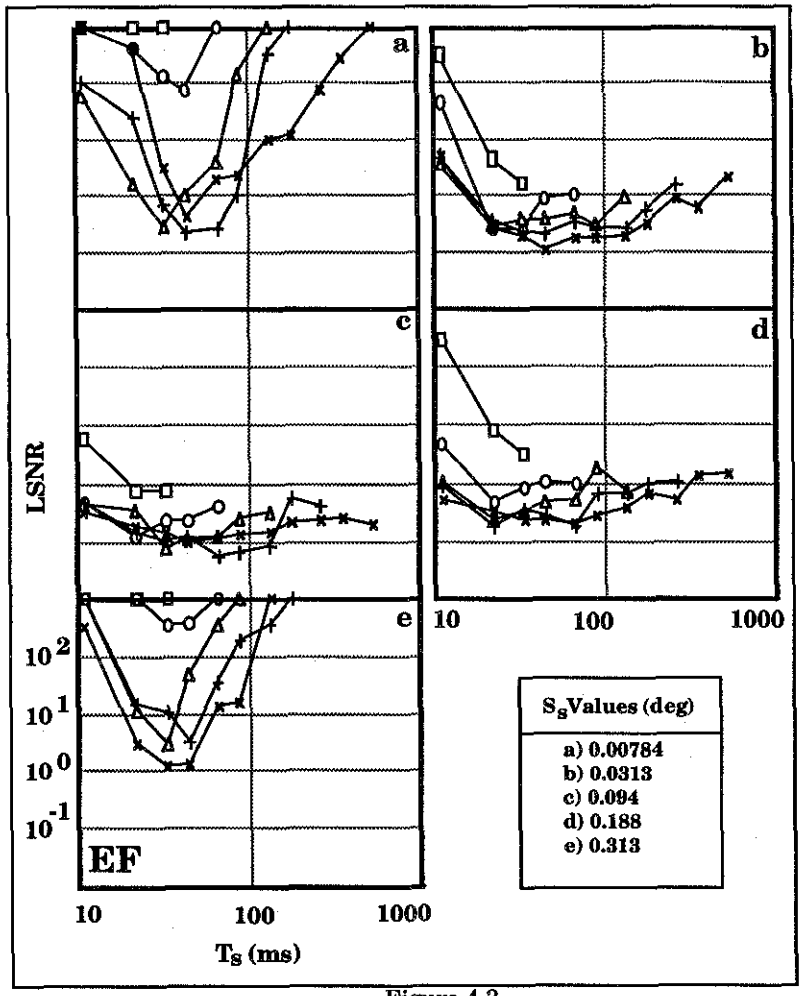


Figure 4.3

The raw data for subject EF is shown. Each graph displays a set of LSNR threshold curves as a function of T_S. The parameter of the curves within each graph is D_S, and the parameter across the graphs is S_S. Unique symbols are assigned to represent a constant D_S: □ for D_S = 66 ms, ○ for D_S = 133 ms, Δ for D_S = 266 ms, + for D_S = 533 ms, and x for D_S = 1066 ms.

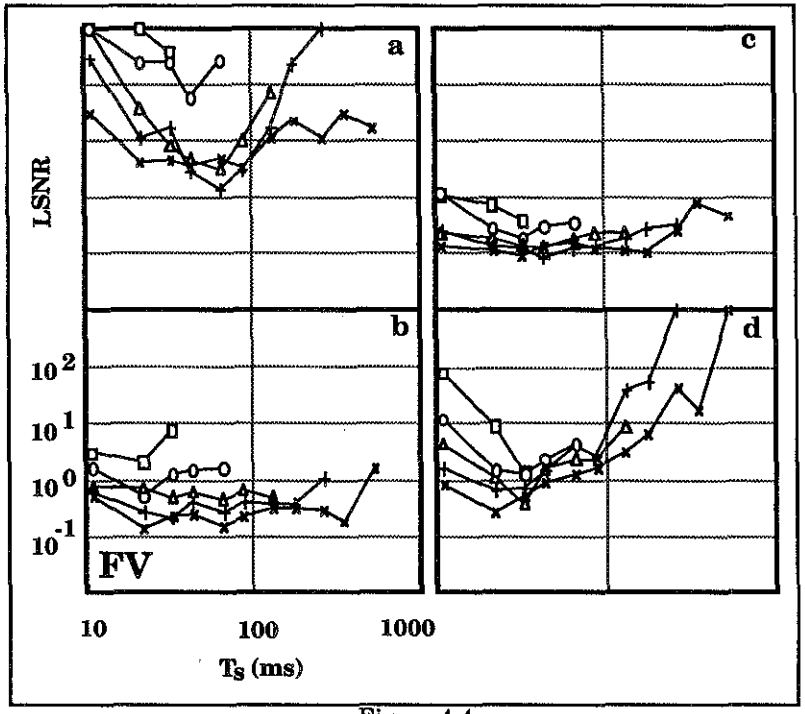


Figure 4.4

Figure 4.4 shows the raw data for subject FV as in figure 4.3.

Both subjects show flattening in the D_S = 1066 ms threshold curve as T_S increases (figures 4.3b-d and 4.4a-c). Note that as T_S is increased for a given D_S value, N_D is correspondingly decreased. The largest T_S for each curve represents N_D = 1. This flattening of the curve represents independence of the threshold from N_D for large T_S values. The constant-D_S curves are roughly parallel for both EF and FV for the middle range of S_S values tested. The curves for the largest and smallest S_S values show a larger variation of behavior. Threshold saturation can be seen by selecting a single T_S value in any of the graphs and observing the behavior of the improvement of threshold. For the middle range of S_S values, saturation begins first for T_S values at and below 33 ms. Threshold saturation did not occur at larger T_S for the D_S values used in this experiment. This behavior appears to represent a threshold saturation time constant that is a function of T_S. In the remainder of the paper we will refer to the threshold saturation time constant as τ_{sat}, a factor that is implicitly dependent on

T_S. The data of subject FV reflect the same properties as seen in previous experiments. FV's data show a single sensitivity optimum below T_S = 50 ms for small S_S (figure 4.4a). As S_S increases, that the range of optimum sensitivity expands. The thresholds of subject EF also reflect the behavior shown in the previous experiments. There is a wide optimum in the low and middle S_S range (figure 4.3b-d), and a single optimum above T_S = 50 ms for the largest measured S_S value (figure 4.3e). The tuning optima in the D_S = 1066 ms curves (for the middle S_S values) are less obvious, perhaps due to saturation of the threshold curves with increasing D_S.

4.4 Temporal Contrast Sensitivity Experiments

The iso-LSNR curves on the $<T_S,N_D>$ threshold surface in figure 4.2a show the influence of a loss of temporal contrast sensitivity at high frame rates. Temporal contrast sensitivity generally decreases as frame rates increase beyond 20 or 30 Hz. This sensitivity loss is indicated by an approximately N_D-independent decrease of sensitivity (increase of threshold) as frame rate is increased (T_S is decreased). Standard temporal contrast sensitivity measurements are generally made by smoothly modulating the contrast of luminance spots or sinusoidal luminance gratings over time. Our stimulus contains a wide range of spatial frequencies and individual dot luminance values are modulated (non-uniformly across the spatial pattern) in an approximate temporal square wave, so estimating temporal contrast sensitivity for our stimulus using the extant literature (see Van de Grind and Grüsser, 1973, or Kelly & Burbeck, 1984, for reviews) is difficult. The following experiments were therefore designed to provide a measure of temporal contrast sensitivity specifically for our stimulus paradigm.

4.4.1 Detection Thresholds

This experiment was designed following standard temporal contrast threshold measurement methods but used a target stimulus similar to those used in our directional LSNR threshold experiments. A random dot pattern was presented on either the left or right half of the two degree square display area and was one degree in width and two degrees in height. The other half of the display was a flat luminance profile with the same mean luminance as the target random dot pattern. The target random dot pattern was refreshed with a spatio-temporally uncorrelated random dot pattern at the same frame rates as used in the directional LSNR threshold experiments presented above. This can be thought of as setting dot life time to a single image frame, or S_S = K with K being larger than the width of the display. The task of the observer was to select the side of the display upon which the noise motion appeared. Threshold measurement proceeded in the same fashion as the LSNR experiments but with average stimulus contrast as the threshold variable (see Methods). Threshold curves for constant D_S were measured for two observers and are shown in figure 4.5. The thresholds were taken for D_S values of 66 ms, 133 ms, 266 ms, 533 ms, and 1066 ms as in the LSNR experiment presented above.

The detection curves in figure 4.5 are very similar in shape to the directional LSNR threshold curves in figures 4.3c and 4.4c for T_S less than 100 ms. Both sets of curves include similar τ_{sat} variation with T_S: τ_{sat} appears to be smaller for small T_S values than for larger T_S values. However the curves in figure 4.5, unlike the LSNR threshold curves, do not turn upwards as T_S increases. Instead they appear to either converge to some minimum threshold v
alue or continue to decrease with viewing time. We suspect that this differing behavior corresponds to that seen in other temporal contrast threshold experiments. It is generally attributed to the change-over of the detection mechanism from motion or flicker detection to form (in this case, random dot pattern) detection. Comparison of the curve shape to the LSNR threshold curves would seem to indicate that the changeover occurs somewhere between T_S = 100 ms and T_S = 200 ms (5-10 Hz). Although the LSNR and contrast threshold curve shapes are similar for T_S values less than 100 ms, we would like stronger evidence that the temporal contrast sensitivity determines the lower T_S shape of the LSNR curves.

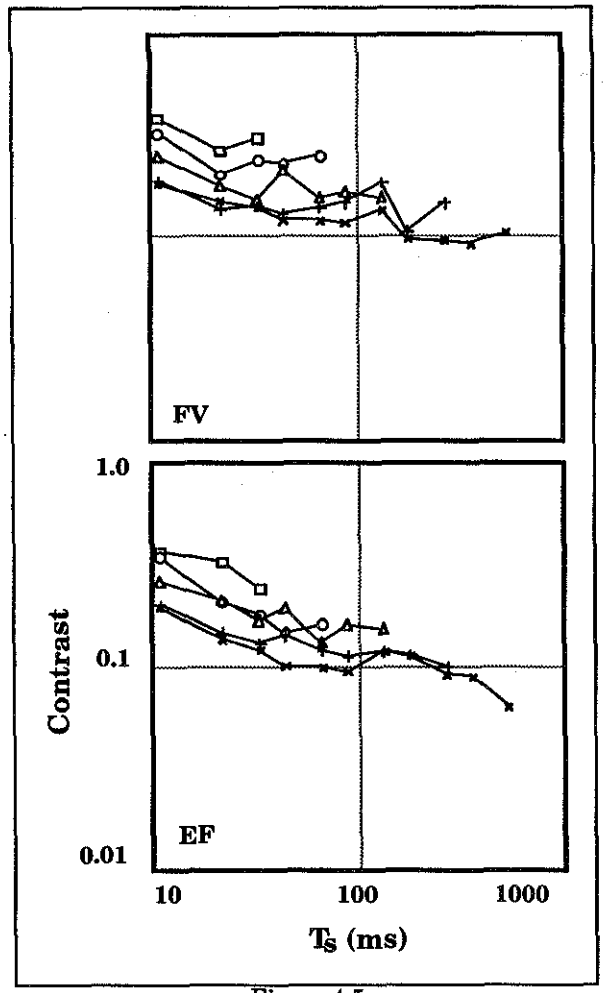


Figure 4.5
Non-directional, noise-motion detection threshold values were measured using average stimulus contrast as the threshold parameter. These curves were measured for two subjects at the same T_S values size as in figures 4.3 and 4.4. The curves were measured with D_S as a parameter across the curves in the graph. The D_S value for each curve is indicated by the same symbols as in figures 4.3 and 4.4.

4.4.2 Directional Contrast Thresholds

Directional contrast thresholds were measured in order to provide a connection between the directional LSNR thresholds and the detection thresholds. In this experiment we used the same stimulus as in the LSNR threshold curves but use average stimulus contrast as the threshold variable. Figure 4.6 shows threshold results for two subjects. Curves were taken for constant-D_S values of 66 ms, 133 ms, 266 ms, 533 ms, and 1066 ms as in the previous experiments. The spatial displacement size of the coherent motion was 0.094 degrees. The threshold measurement procedure was the same as the LSNR threshold curves except that average contrast was the threshold measurement variable.

These curves show the same behavior as the directional LSNR threshold curves in figures 4.3 and 4.4. At the lower T_S values the directional contrast threshold curves show approximately the same absolute values as the detection thresholds in figure 4.5. The curves are essentially the same until T_S gets larger than 133 ms, at which point the detection threshold curves (figure 4.5) continue to decrease towards a value that is lower than that for a directional threshold (figure 4.6). The combination of these two experiments would seem to indicate that temporal contrast sensitivity is a good predictor of the LSNR threshold curve behavior for T_S values less than about 100 ms, at least for an S_S value of 0.094 degrees.

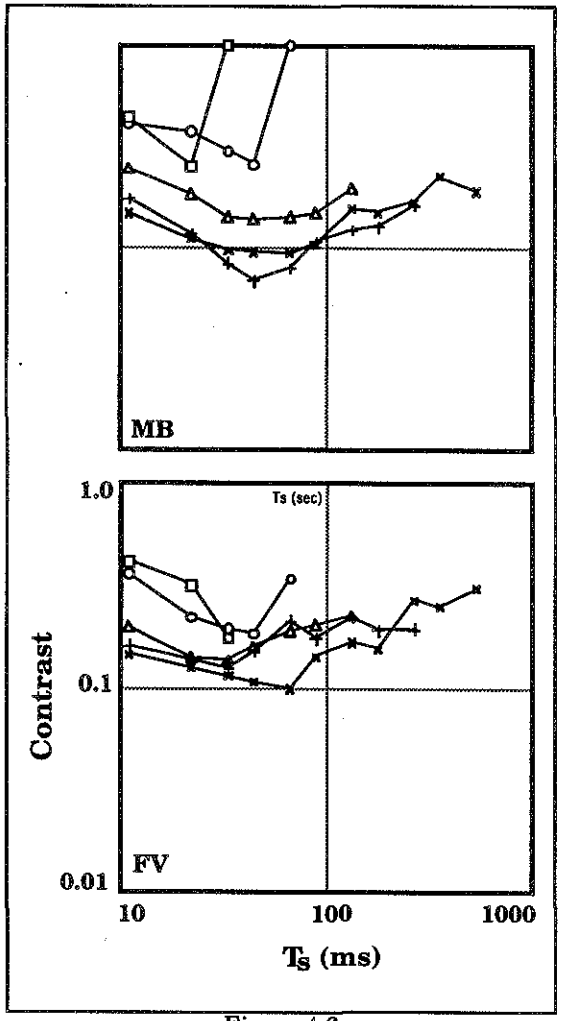


Figure 4.6
Directional motion detection threshold values were measured using average stimulus contrast as the threshold parameter. These curves were measured for two subjects at the same S_S value as in figures 4.3c and 4.4c. The curves were measured using the same T_S values as in figures 4.3, 4.4 and 4.5, with D_S as a parameter across the curves in the graph. The D_S value for each curve is indicated by the same symbols as in figures 4.3 and 4.4.

4.4.3 Discussion

A more direct link between the LSNR and contrast thresholds for directional motion detection can be made if we examine the LSNR method in detail. As described by van Doorn and Koenderink (1982a, 1982b), the stimulus contrast and LSNR in the LSNR threshold measurement method are determined by

$$C_{stimulus} = \sqrt{C^2_{signal} + C^2_{noise}} \tag{4.2}$$

$$LSNR = \frac{C^2_{signal}}{C^2_{noise}} \tag{4.3}$$

where C_{signal} and C_{noise} are the root-mean-square (rms) contrast of the signal and noise patterns, respectively, and $C_{stimulus}$ is the average contrast of the whole stimulus. LSNR is then defined as the ratio of the square of the signal and noise contrasts (the ratio of the power of the two signals). During threshold measurement $C_{stimulus}$ is held constant and the LSNR is determined by equations 2 and 3. We can therefore calculate C_{signal} and C_{noise} for any measured LSNR threshold value and re-plot the curves on a contrast scale. The relationship is

$$C_{signal} = C_{stimulus} \cdot \sqrt{\frac{LSNR}{LSNR + 1}} \tag{4.4}$$

$$C_{noise} = C_{stimulus} \cdot \sqrt{\frac{1}{LSNR + 1}} \tag{4.5}$$

We will refer to the transformed threshold curves as C_{signal} or C_{noise} threshold curves. The C_{signal} transformation was performed for the S_S = 0.094 degree LSNR threshold curves for subject FV (figure 4.4c); the result is shown together with the subject's corresponding directional contrast threshold curves (figure 4.6a) in figure 4.7. The difference between the C_{signal} and directional contrast threshold values was about 0.1 contrast units for all data points. The LSNR thresholds used to calculate the C_{signal} curve were all between about 0.1 and 1.0, and most were between

LSNR values of 0.1 and 0.3. The luminance noise strength in the motion signal was approximately constant (values between about 0.5 and 0.65). The required shift of 0.1 contrast units represents the influence of the noise in the motion signal. The constant shift of 0.1 contrast units is consistent with the approximately constant strength of the noise in the signal. The C_{signal} threshold curves were therefore shifted down by a value of 0.1 contrast units. The constant shift brought the two curve sets into good correspondence.

There is a good correspondence between the shapes and absolute values (modulo the additive constant of 0.1 described) of the contrast and C_{signal} threshold curves. The staircase procedure used the same step sizes in both threshold experiments. This results in a varying step size for the C_{signal} threshold curves relative to the LSNR step size. Nevertheless, consideration of figures 4.5, 4.6 and 4.7 leads us to conclude that temporal contrast sensitivity is indeed a good predictor of directional LSNR threshold curve shape for T_S values lower than about 133 ms. The shape of the directional LSNR threshold curves in figures 4.2 and 4.3 for T_S values larger than 133 ms must be determined at least in part by some other mechanism than temporal contrast sensitivity.

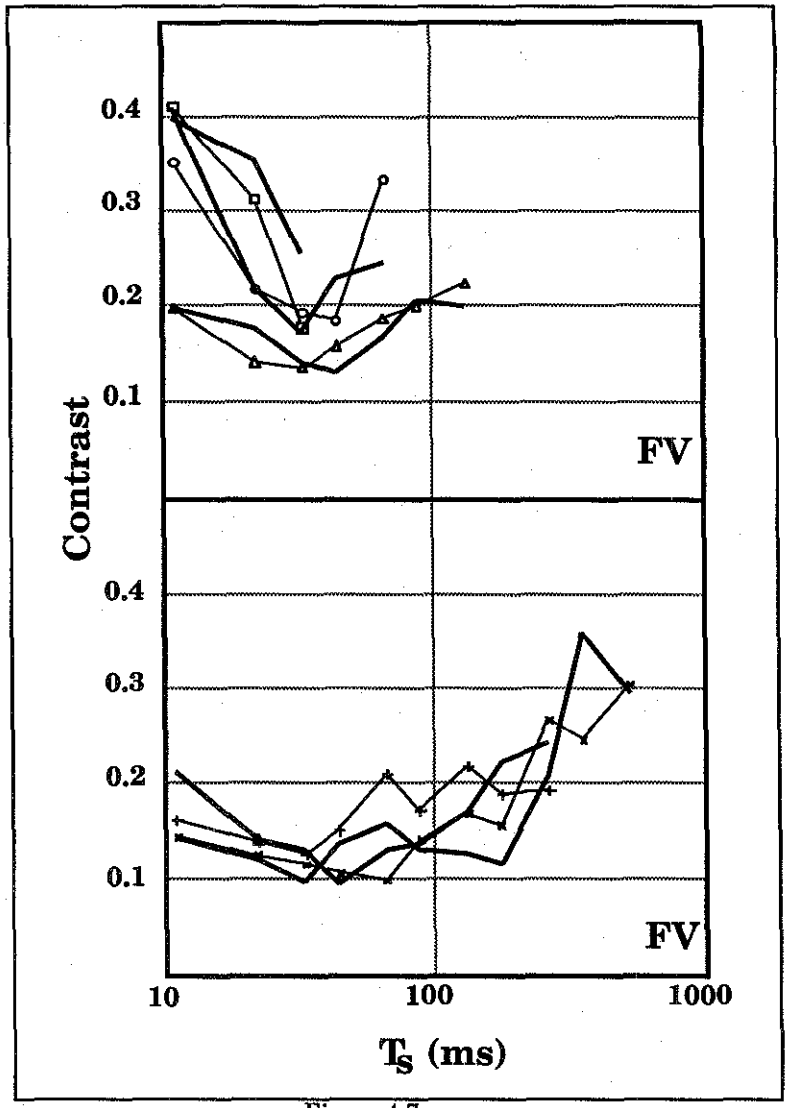


Figure 4.7
Comparison of the contrast threshold curves (figure 4.6) with the C_{signal} curves (calculated from figure 4.4c) for subject FV. Lines with symbols are contrast threshold curves. Heavy lines without symbols are C_{signal} curves. Curves of the same length represent the same constant-D_S value. The symbols have the same meaning as in figures 4.3 and 4.4.

4.5 Temporal Integration Model Extension

We now discuss the formulation of a more precise model for temporal integration of motion information. Our previous high-level model for temporal integration (Fredericksen et. al, submitted; chapter 3) assumed that the directional LSNR threshold is inversely proportional to D_S and includes the influence of temporal contrast sensitivity loss. Figure 4.2b shows a contour diagram of that model threshold surface, defined as

$$\text{LSNR}_{\text{estimate}} = \frac{K}{D_S(T_S, N_D) \cdot CS(T_S)} \tag{4.6}$$

for some arbitrary constant K, where D_S is defined as above and the effect of contrast sensitivity is described by

$$CS(T_S) = 10^{-e^{(A \cdot \log(T_S)+B)}} \tag{4.7}$$

(Fredericksen et. al, submitted; chapter 3). This model provides a good first order representation of the behavior of the temporal integration of motion information but has several deficiencies. The model does not include the influence of the populational T_S tuning other than a rough estimate of contrast sensitivity. Nor does it account for the T_S-dependent saturation times shown by the current threshold data, or threshold saturation in general. The psychophysical experiments presented above and neurophysiological information from the literature are useful in refining this model.

4.5.1 Threshold saturation and τ_{sat}

The leaky integration characteristics of neural cells is well known, and leaky integrator components have previously been used in a model of alpha-stripe[5] generation during the visual perception of motion in random dot patterns (Adler, Bock, and Grüsser, 1981). The behavior of motion-detecting cells in the visual pathway when subjected to apparent motion stimuli can help us understand the leaky integration mechanism. Apparent motion stimuli provide discrete, pulsatile visual motion information. We know that directionally selective cells in V1 and MT of macaque show analogous discrete responses to stroboscopic bar motion (Mikami, Newsome and Wurtz, 1986a, 1986b). As the temporal delay (short T_S plus ISI) between bar presentations in those experiments increases, the individual bursts of activity become individually distinguishable. At shorter temporal delays the neural response pulses merge, a condition approximating response to a "real" motion stimulus. Neurons in macaque visual area MT also show evidence of temporal integration and response saturation over time (cf. Mikami et al., 1986a, figure 4). That is, they can show an increasing mean firing rate over time, eventually reaching a maximum mean firing rate (cf. Mikami et al., 1986a, figure 4). Our psychophysical threshold measurements are a function of all stimulated motion detectors that contribute to the directional motion percept. If most or all individual detectors in the population contributing to a percept exhibit a similar property, we might expect that property to appear in the psychophysical measurement. We should therefore not be surprised if the leaky integration properties of directionally selective cells along with those cell's pulsatile responses to discrete spatial displacements in the apparent motion stimulus are reflected in our psychophysical measurement. For these reasons the extended model of temporal integration of motion information presented here is based on the leaky integration of the output of hypothetical motion detectors that respond to the manner indicated above.

4.5.2 Temporal contrast sensitivity and populational tuning

Our refined model also addresses the influence of T_S on the temporal integration process. This includes two effects shown in our data: temporal contrast sensitivity and T_S tuning optima. We have provided evidence in this paper that the improvement of the directional LSNR threshold curve is predicted by temporal contrast sensitivity for T_S up to about 133 ms. We model more precisely the influence of temporal contrast sensitivity by using the contrast threshold curves above to modify the magnitude of the motion detector output pulse. The directional contrast threshold data for $T_S < 100$ ms for subject EF was normalized to the maximum contrast value ($T_S = 89$

[5] If a subject views random dot pattern apparent motion with a fixed point of gaze, then for some frame rates a stationary striation of the random dot pattern is seen perpendicular to the direction of motion. These striations are called alpha-stripes and represent an internally generated modulation of the percept.

ms) and plotted on a log-log scale (see figure 4.8) as sensitivity (the inverse of the threshold value, divided by the threshold value at T_S = 89 ms). A second degree polynomial was fit to this contrast sensitivity curve on a log-log scale as described by equation (4.8). The contrast sensitivity curve therefore increased smoothly to a value of 1.0 at T_S = 89 ms. All T_S values above 89 ms were assigned a contrast sensitivity of 1.0.The magnitude of the current pulses input to the leaky integrator circuit were assigned according to this contrast sensitivity curve.

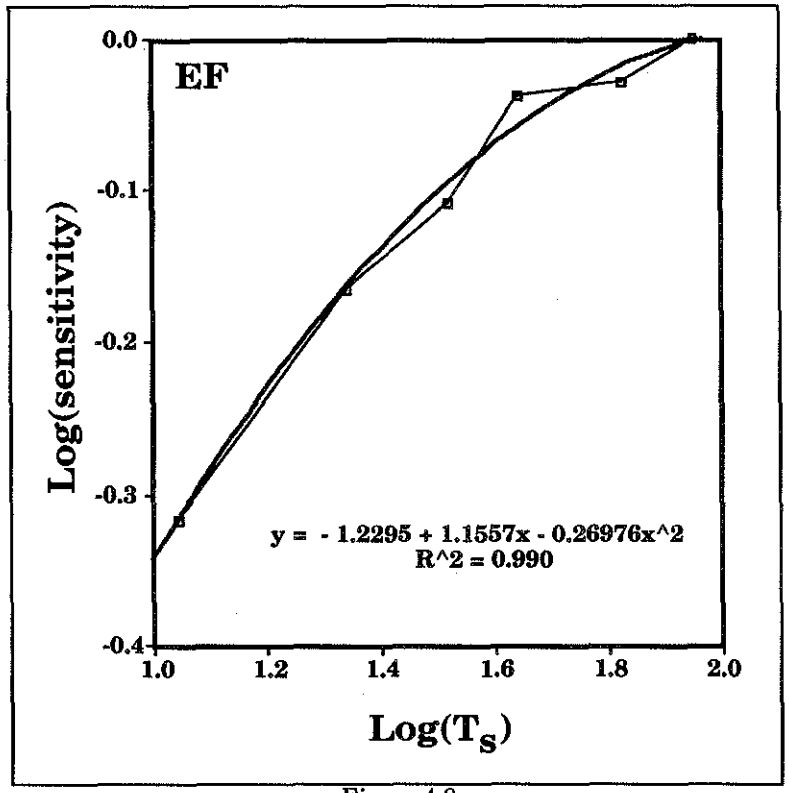


Figure 4.8
The calculated contrast sensitivity data and the fitted curve used in the leaky integrator simulation.

$$CS(T_S) = 10^{A+B\cdot\log(T_S)+C\cdot\log^2(T_S)} \tag{4.8}$$

There is support for variation of motion detector T_S tuning and the dependence of that tuning on the value of S_S in the neurophysiology literature (Mikami et al., 1986a, 1986b) and the psychophysical literature (Van de Grind, et. al., 1986). These properties seem to be present independent of the value of D_S, indicating that they may be independent of the temporal integration mechanism. Inclusion of these properties in the model requires hypothesis of a mechanism. We defer the topic of motion detector tuning to the general discussion and do not attempt to model it.

4.5.3 Stimulus Model

The pulsatile output of direction selective cells in visual cortex (cf. Mikami et. al., 1986, figure 4) is a direct response to pulses of motion information fed to them. The aggregated output of our hypothetical motion detectors responding to an apparent motion stimulus was therefore modeled as a sequence of current pulses. Each pulse corresponds to a single displacement of the apparent motion stimulus. The pulses were fed to the leaky integrator circuit spaced in time according to T_S in the stimulus. The current input wave form was a superposition of a number of input pulses with an abstract form of a neural cell's impulse response function (Koenderink, 1988). The form of the pulse was designed to provide causality in a temporal multiresolution filter system. The pulse form is derived through a particular re-mapping of the time axis for all past time up to the present moment (that is, ignoring the future). The input pulse shape, defined by

$$\text{pulse}(t) = u(t)\cdot e^{\frac{-\ln\left(\frac{t}{\tau}\right)}{4\sigma}} \tag{4.9}$$

can be thought of as representing current influx to a neuron due to neurotransmitter released at synaptic junctions, where each pulse represents a single discrete spatial displacement in the apparent motion stimulus. The two parameters of the function determine its shape and can be interpreted as temporal resolution and scale. The resolution parameter τ determines the position of the pulse peak and is analogous to localization of an event (measurement) in time. The scale parameter σ determines the width of the pulse and is analogous to duration of an event

(measurement) in time. The peak magnitude of each pulse was selected from a contrast tuning curve estimated from our psychophysical data (see above).

4.6 Temporal Integration Model Simulation

We will first describe the simulation and then discuss the correspondences between the model characteristics and the psychophysical threshold data. A leaky integrator circuit was simulated using electrical circuit simulation software (see appendix A). The electric circuit was a resistor and capacitor in parallel to ground, with a current input source[6]. The resistive and capacitive properties of neurons are well known, and the resistor-capacitor circuit is an accepted model of those properties. The output voltage of the circuit (i.e. across either the resistor or capacitor) can be thought of as the filtering of a motion detector's response to the visual stimulus with the goal of improving the detector's sensitivity. The sensitivity improvement due to the leaky integrator is a function of the maximum output magnitude that it reaches in response to any sequence of input images fed to the motion detector.

The resistor-capacitor circuit is shown in figure 4.9a. The contrast sensitivity curve shape is shown in figure 4.9b . Examples of input pulse train shapes, modified by the contrast sensitivity tuning curve, and the corresponding integrator output curves are shown in figure 4.9c. The parameters of the leaky integration circuit and the input current pulse shape were selected so that the simulation output approximated the middle S_s value threshold curves in figures 4.3 and 4.4.

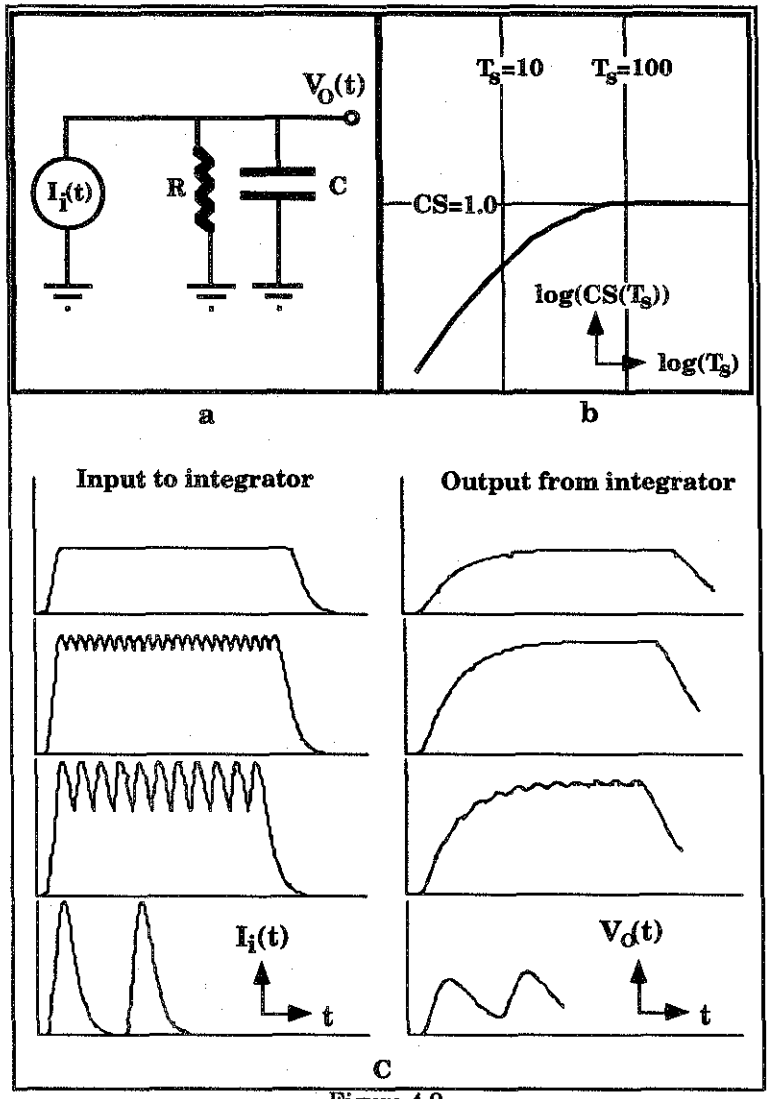


Figure 4.9
(a) The circuit used in the leaky integrator simulation. (b) The contrast sensitivity curve used as the modifier of the pulse height, based on the contrast sensitivity data of subject EF. (c) Example current input pulse curves are shown on the left while corresponding output curves are shown on the right.

[6] It is often possible to calculate a closed form solution for the transient response of a simple resistor-capacitor circuit. However it is more convenient to use a numerical solution method if manipulation of major properties of the circuit is desired (e.g. component count and configuration, or the input current pulse shape).

4.6.1 Threshold Calculation

We want to more closely compare our simulation results to our psychophysical threshold results. To do so, we must think carefully about how the output of our leaky integrator model is related to the populational response of the motion detectors to apparent motion. (The reader should keep in mind the high-level model in figure 4.1 throughout this section.) Our threshold measurements are stated in terms of the spatial contrasts of the input signal and noise, but the output of our model is a voltage level. The goal, then, is to relate this voltage level to an equivalent input pattern contrast. We first examine the influence of spatial pattern contrast on the leaky-integrated response of the (biological or model) motion detector array by examining the model structure and physiological literature. We then discuss how the output voltage level of the leaky integrator is related to the populational response of the motion detector array, and hence a hypothetical (or equivalent) input pattern contrast. The equivalent model contrast can then be used to compute an equivalent LSNR threshold.

4.6.1.1 Spatial pattern contrast and populational response

The populational response is a function of a number of parameters, including spatio-temporal structure of the stimulus (spatial structure Ψ, frame rate $1/T_s$), stimulus duration (D_s), mean luminance (**L**), target stimulus contrast (**C**), stimulus position and size in the visual field, etc.

$$R_{pop} = f(\Psi, T_s, D_s, L, C, \text{etc...}) \tag{4.10}$$

Equation (4.10) will be used below as a basis for describing the influence of temporal integration on the final perceptual threshold.

The next step is to decide how the motion detectors respond to motion input (and hence the output of the leaky integrator). Visual motion detectors are commonly modeled as either bi-local detectors (Reichardt, 1961) with a multiplicative correlator component (e.g. see figure 4.1) or as spatio-temporal energy detectors (Adelson and Bergen, 1985) that have a very similar architecture. In addition, directionally selective neurons in cat visual cortex have been shown to have characteristics consistent with the spatio-temporal energy detector mechanism (Emerson, Bergen and Adelson, 1992). We are not championing either model here. Both mechanisms compute motion information using the output of spatial filters covering visual space, and have a preferred spatial displacement size and velocity (spatio-temporal structure of the input). The important point is that both detector mechanisms produce output values that are proportional to the square of the front-end filter response magnitude .

The input to the multiplicative correlator comes from the front-end receptive fields (filters) of the motion detector. In the primate visual system these front-end filters are likely to correspond to the receptive fields of lateral geniculate nucleus cells. A model for the luminance contrast response (firing rate) of retinal ganglion and lateral geniculate nucleus (LGN) cells in cat (Robson, 1980) is

$$R_{lgn} = k \cdot \log\left(1 + \frac{C}{C_0}\right) \tag{4.11}$$

where R_{lgn} is the firing rate of the cell, **C** is the luminance contrast of the input pattern, and **k** and C_0 are constants. C_0 essentially determines the stimulus contrast level above which the response of the cell (to stimulus contrast) is no longer linear. The populational response of the LGN cells to luminance contrast is an aggregate of these responses, and the value of C_0 apparently varies widely across the cell population (Robson, 1980). The contrast response for low contrast values (up to the populational value of C_0) may therefore be modeled by a linear function. The linear contrast response of the front-end filters is followed by the detector's squaring operation. The motion detector response is therefore proportional to the square of the input pattern contrast, C^2 (again, for values of **C** below C_0).

The last step in the processing sequence is the leaky integration of the output of the detector population. Leaky integration is a linear function, and the output is therefore also proportional to C^2. If we assume that the contrast response function is separable from f() in equation (4.10), then we have

$$R_{pop} = C^2 \cdot g(\Psi, T_s, D_s, L, \text{etc...}) \tag{4.12}$$

where g() represents the remainder of f() after separating out the contrast function and includes the influence of temporal integration.

4.6.1.2 Correspondence of model output to response threshold

By convention, the measured psychophysical threshold is considered as inversely proportional to populational response magnitude. Moreover, the response of the motion detecting system must be greater than some internal noise level before the target stimulus is perceived. We can call the required populational response R_{min} (R_{min} is a measure of the internal noise) and use equation (4.12) to write

$$C^2_{min} = \frac{R_{min}}{g(\Psi, T_s, D_s, L, etc...)} \tag{4.13}$$

where $\mathbf{C_{min}}$ is the contrast required to reach the threshold-response level. The function g() represented the leaky-integrated response of our motion detector array. The peak value of g() determines the minimum threshold over the leaky integration period. In terms of our leaky-integrator model in figure 4.9, g() is the voltage level of the circuit. We can therefore state the relationship of the leaky-integrator output to the luminance contrast of a hypothetical pattern input as

$$\mathbf{C_m^2} \propto \frac{1}{\max_t[V_0(t)]} \tag{4.14}$$

where $\mathbf{C_m}$ is the "luminance" contrast value for the hypothetical input pattern.

4.6.1.3 Calculating an equivalent model LSNR

Equation (4.14) provides a method for relating the model output to the contrast of a hypothetical motion input pattern. One final step is required because the leaky-integrator simulation does not directly address the influence of noise in the motion signal. The difference of 0.1 contrast units between the directional contrast thresholds and $\mathbf{C_{signal}}$ thresholds indicates that the noise in the motion stimulus has some influence on the psychophysical threshold level. Moreover, the psychophysical thresholds are stated in terms of a signal to noise ratio. Fortunately we can bypass the influence of noise by assuming that the leaky-integrator output was produced in the presence of noise rather than in a noise-free situation (adding noise to the input increases the value of R_{min} in equation (4.13)). We then use equation (4.4) to calculate a model pseudo-LSNR value from $\mathbf{C_m}$ as

$$LSNR_m = \frac{\mathbf{C_m^2}}{\mathbf{C^2_{stimulus}} - \mathbf{C_m^2}} \tag{4.15}$$

where $\mathbf{C_{stimulus}}$ is the average stimulus contrast in the psychophysical threshold measurements. In the psychophysical thresholds, $\mathbf{C_m}$ is equivalent to $\mathbf{C_{signal}}$.

4.6.1.4 Absolute threshold

In general, our perceptual thresholds are influenced by all parameters of the stimulus. In our psychophysical experiments we fix as many parameters as possible and measure a threshold by manipulating the LSNR while keeping $\mathbf{C_{stimulus}}$ constant. Equation (4.15) indicates that $\mathbf{C_{stimulus}}$ determines an absolute threshold level in the same manner for both the psychophysical experiments and the leaky-integrator model simulation. Inspection of equation (4.15) also shows that, due to the non-linear relationship between $\mathbf{C_m}$ ($\mathbf{C_{signal}}$) and $LSNR_m$ (LSNR), as $\mathbf{C_m}$ ($\mathbf{C_{signal}}$) approaches $\mathbf{C_{stimulus}}$, $LSNR_m$ (LSNR) increases very quickly. The strongest coherent motion signal that we can produce in the psychophysical experiment is a noiseless one in which $\mathbf{C_{signal}} = \mathbf{C_{stimulus}}$. Consequently, the value of $\mathbf{C_{stimulus}}$ will also determine the visible ranges of the other stimulus parameters when using the LSNR threshold method. Use of equation (4.15) to plot the simulation output incorporates the influence of the absolute threshold behavior into the leaky integrator model.

4.6.2 Results and Discussion

We begin by discussing the relationship of the model parameters to the simulation output curves shown in figure 4.10. The threshold optimum of $LSNR_m$ curve is a result of temporal contrast sensitivity improvement as T_s increases and the reduced temporal integration caused by increasing

τ_{sat}. Temporal contrast sensitivity accounts for the decreasing threshold as T_s increases from 11 to around 44 ms. The upturn of the $LSNR_m$ threshold curve around T_s = 133 ms is due to increasing τ_{sat}. The flattening of the $LSNR_m$ curve as T_s increases further represents the increasing independence of the input pulses: when they are far enough apart, they are no longer summed by the leaky integrator. The width of the local threshold optimum is a function of both the input pulse shape and the intrinsic time constant, τ_{rc}, of the leaky integrator. In the model τ_{sat} is an effective saturation time constant produced by the combination of τ_{rc} and the input pulse shape and magnitude. We use the term "effective" because the rate at which the output improves across input pulses is not the same as for a unit step input pulse (determined completely by τ_{rc}). As the pulse spacing increases, the integrator has more and more time to discharge between input pulses and thus τ_{sat} increases. τ_{sat} approaches τ_{rc} as the pulses are spaced closer together (as T_s decreases).

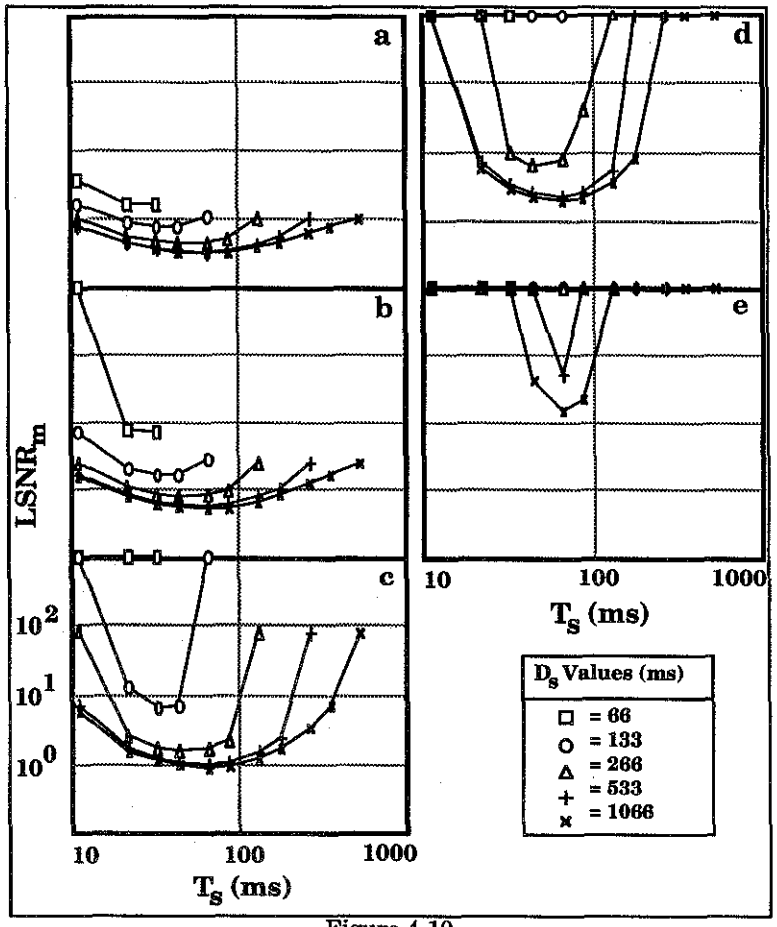


Figure 4.10
Model threshold output approximating the threshold curves shown in figures 4.3 and 4.4. The curves are plotted in the same manner as the psychophysical data in figures 4.3 and 4.4. The method for calculating the model threshold is explained in the text. The D_s value for each curve is indicated by the same symbols as in figures 4.3 and 4.4. The magnitude of the current pulse decreases from 4.10a to 4.10e.

4.6.3 Comparison of LSNR and $LSNR_m$ Thresholds

We do not calculate a quantitative measure of similarity between the LSNRm curves in figure 4.10 and the directional LSNR threshold curves, but visual inspection of the $LSNR_m$ threshold curves should be adequate to

convince us that the leaky integrator model captures important properties of the psychophysical measurements. The sequence of figures from a to e in figure 4.10 represent decreasing visibility of the stimulus. The sequence of figures 4.10a-e are equivalent to moving from a middle S_S value in figures 4.3 and 4.4 (e.g., figure 4.4c) toward either larger or smaller S_S values in those same figures (e.g., figure 4.3a, or 4.3e). The value of τ_{sat} in both the simulation (figure 4.10) and the psychophysical data (middle S_S values in figures 4.3 and 4.4) depends on T_S. We note that the T_S tuning optima in the psychophysical curves for shorter D_S are often sharper than those in the simulated curves. A possible explanation of this is given in the general discussion below.

One mechanism of the model that is not as clear in figure 4.10 is the influence of response thresholding. To more clearly illustrate the T_S dependent saturation levels and time constants, and the influence of a response threshold mechanism, we re-plot in figures 4.11 and 4.12 the data from figure 4.3 (subject EF) and figure 4.10 (the model) as constant T_S curves (the axis into the page for figures 4.3 and 4.4) with D_S as the independent variable. The psychophysical data in figures 4.11a-4.11e are shown in order of increasing S_S. The simulation data in figures 4.12a-4.12e are shown in order of decreasing current magnitude. Inspection of the graphs reveals similar behavior across them: shorter τ_{sat} values are correlated with shorter T_S values with progressive loss of stimulus visibility at the extremes of S_S. We point out here an important behavior of these curves. At the extremes of the perceivable S_S ranges we see a loss of visibility, and a variation of the values of τ_{sat} across S_S that at first glance might be interpreted as S_S dependence. However we would like to propose a different explanation based on the populational nature of the psychophysical percept.

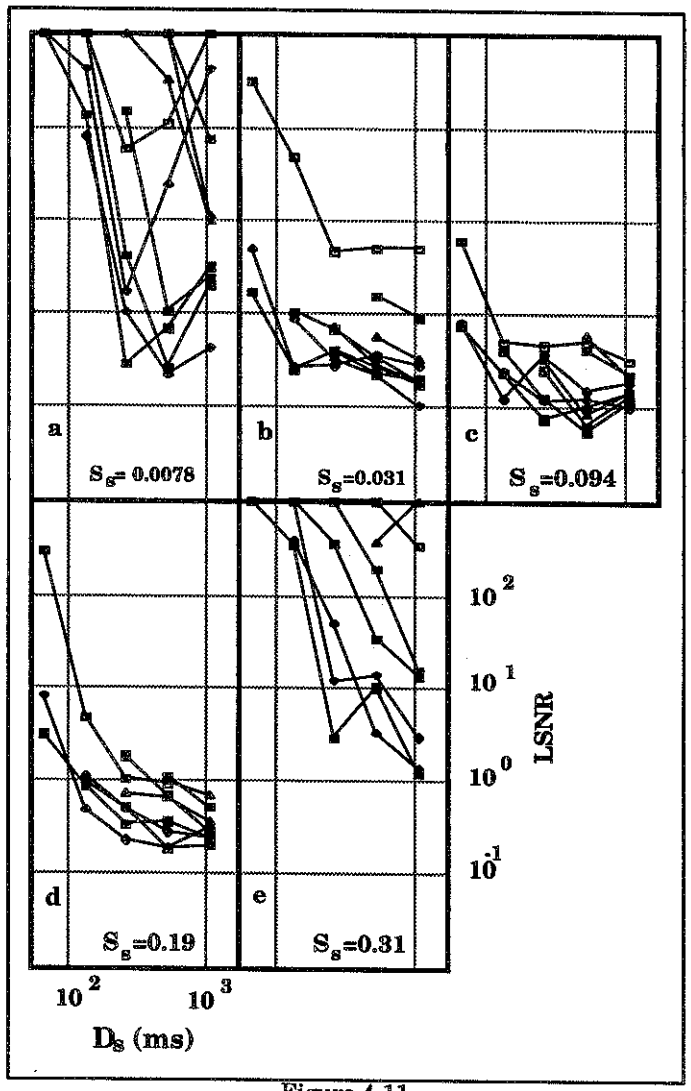


Figure 4.11
The data from figure 4.3 are re-plotted as constant-T_S curves as a function of D_S. The parameter across the graphs is S_S.

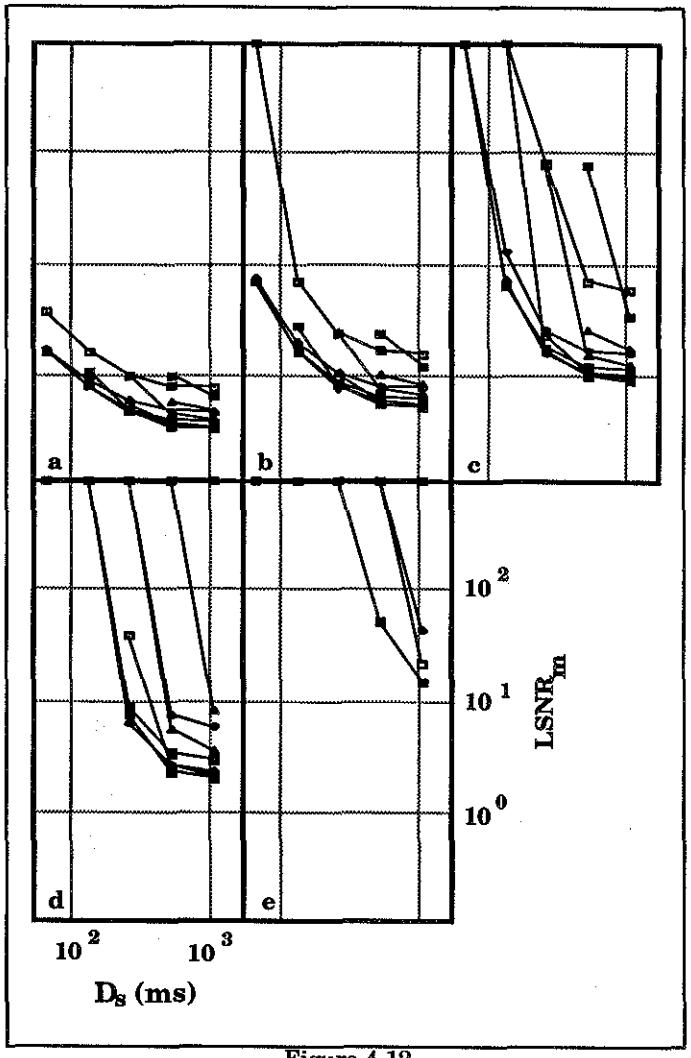


Figure 4.12
The data from figure 4.10 are re-plotted as in figure 4.11. The parameter across the graphs (a to e) is increasing pulse magnitude.

In a previous paper (Fredericksen et. al, submitted; chapter 3) we showed evidence of spatial summation in our stimulus, as has been previously shown in random dot apparent motion stimuli with $N_L >> 1$ (Van Doorn and Koenderink, 1984). The psychophysical threshold changes with the changing height of the random dot stimulus, asymptotically following the relation predicted by probabilistic summation of motion detector outputs (Van Doorn and Koenderink, 1984). This spatial summation property was used to remove the effects of threshold saturation on the $<T_S,N_D>$ plane shown in figure 4.2a (the flattened threshold regions at large $<T_S,N_D>$ values) through reduction the height of the stimulus. A curious result of this test, however, was that a pair of 0.5 degree high, 2 degree wide strips of motion stimulus above and below the fixation point produced a greater rise in threshold values than did a 0.5 degree, 2 degree wide strip of motion stimulus in the center of the display . In both cases the remainder of the 2x2 degree display area was filled with a flat luminance profile with the same mean luminance as the motion stimulus.

This effect is curious because, counter to the probabilistic summation relationship, a greater amount of stimulus area resulted in higher (worse) threshold values. The upper and lower sides of the single central strip were flanked by mean luminance regions of display. In the two strip display, however, the upper edge of the upper strip and the lower edge of the lower strip were at the edge of the usable CRT display area and were flanked by dark regions (the portion of the CRT outside the display area) resulting in a high spatial contrast. We hypothesized that this spatial contrast at the edge of the stimulus may inhibit the percept but also proposed that this effect may reflect the influence of a non-uniform distribution (across the 2x2 degree display region) of motion detectors responding to that motion stimulus. There is a good deal of evidence for variation of the extremes of visible displacement sizes (and other parameters of the motion detecting machinery) with eccentricity in the visual field (Baker and Braddick, 1985a; Van de Grind et. al., 1986; Fredericksen et. al., in press; chapter 2). This variation of parameters with eccentricity in the visual field represents a non-uniform distribution of spatial displacement size detection. It would be surprising if such non-uniform detector distributions do not hold for detectors responding to other displacement sizes as well. We propose that the combination of detectors over a non-uniform detector distribution in the visual field, and a response threshold in the motion information calculation account for the variation of τ_{sat} with S_S.

This hypothesis can be tested in the model. The combination of detector outputs across the population corresponds to an increase in the magnitude of the current pulse in the simulation. The amount of pulse magnitude change with S_S depends on the number of detectors responding to the stimulus and the output combination function. If we assume that the combination function is monotonically increasing with the number of participating detectors (and assuming roughly equivalent individual detector responses), then the data in figure 4.10 indicate that more detectors respond to central S_S values than to the extreme S_S values. The number of hypothetical detectors responding to the stimulus is small for small S_S, increases as S_S is increased, then decreases again. Such a distribution is consistent with our explanation for the dependence of the limits of perceivable spatial displacement given in a previous publication (Fredericksen et. al, in press; chapter 2).

In our model, changing the number of combined detectors is equivalent to changing the current pulse magnitude. This results in a shift of the $LSNR_m$ curves up the logarithmic z- axis (addition of a constant). The response threshold included in the $LSNR_m$ then produces the τ_{sat} variation with S_S. The influence of the response threshold is greater for higher values of $LSNR_m$. The result is a differential vertical stretching of portions of the $LSNR_m$ curves that represent hypothetical pattern contrasts near a value of $C_{stimulus}$. This results in an apparent variation of τ_{sat} with S_S in the LSNR threshold curves. The $LSNR_m$ curves in figure 4.12 were made by calculating the same data points as in figure 4.10 for different maximum pulse magnitudes (equivalent to changing the maximum sensitivity in the $CS(T_S)$ function from 1 to some other value) and plotting them as in figure 4.11. Once again visual inspection of the similarities between the psychophysical and model threshold curves would indicate that the model captures an important property of the psychophysical measurements.

4.8 General Discussion

There are several populational properties that we do not try to capture precisely in our model. The directional motion thresholds include the influence of spatial summation, as discussed above, and the variation of T_S tuning with both S_S and eccentricity in the visual field (Fredericksen et al., in press; chapter 2; Fredericksen et. al., submitted; chapter 3). In the experiments presented here the size and position of the stimulus in the visual field was held constant, so spatial summation effects could be neglected. A discussion of spatial summation and its relation to temporal integration will be presented elsewhere (manuscript in preparation).

The psychophysical data presented here and in our previous papers indicated two extremes of detector response: detectors that have a maximum T_S value to which they respond, and detectors that do not (for the range of T_S values measured; testing of longer T_S values may reveal a maximum even for the persistent motion detectors). These temporal tuning characteristics also appear in the neurophysiological results of Mikami, Newsome and Wurtz (1986a, 1986b) when using stroboscopic bar stimuli. We must be careful in comparing their neurophysiological results to the psychophysical results presented here because we used no interstimulus interval while their stroboscopic bar experiment did. Reduction of the strobe rate in those experiments involved extension of the interstimulus interval. In a mechanism like a bi-local detector it is guaranteed that increasing the ISI will eventually silence the detector because the receptive field outputs go to zero for the flat luminance pattern during the ISI (assuming that the receptive field does not respond to unchanging illumination). This effect will occur regardless of the temporal response characteristics of the receptive field components or their response to luminance flashes. If the reduction of the strobe rate (frame rate) is done by increasing the exposure duration of the images (with no ISI, as in our experiments), we can distinguish between detectors that have either transient or sustained receptive fields responses to a spatial luminance pattern. Detectors that have transient-type receptive fields will fall silent as T_S is increased, but detectors with sustained-type receptive fields will continue to respond. The transient/sustained response property in the stroboscopic bar experiment is evidenced by longer or shorter maximum ISI values that should follow the response duration of the receptive fields to the stroboscopic flash. Thus the physiological and psychophysical data both support the existence of a range of detectors with the above transient/sustained response properties.

The sustained/transient behavior of motion detectors is contained in the current pulse shape used to model the detector output in the simulation. The pulse shape is a function constructed so as to have some interesting and important theoretical properties (Koenderink, 1988). If temporal scale and resolution are important measures in the calculation of motion information, the temporal scale and resolution of the receptive field responses should be preserved in the motion detector response: motion detector responses should match the transient/sustained responses of their receptive field components. There is psychophysical and neurophysiological evidence that motion detector receptive field size increases with $\mathbf{S_S}$ (or optimum spatial displacement size in directionally selective cortical cells) (Van de Grind, et al, 1986; Mikami, et. al., 1986), and that temporal response properties are inversely related to $\mathbf{S_S}$ (van de Grind et. al., 1986; Mikami et. al, 1986; Fredericksen et al., in press; chapter 3). These properties may account for the populational tuning curve variation with $\mathbf{S_S}$ that we observe in the psychophysical data (i.e. $\mathbf{S_S}$ dependent $\mathbf{T_S}$ tuning optima). These populational characteristics may also result in the sharp optima in the low $\mathbf{D_S}$ (66 to 266 ms) psychophysical threshold curves that do not occur in the $LSNR_m$ threshold curves. Transient-type responses are generally faster (high temporal resolution) and would contribute more quickly to the leaky integration process. Due to this possible complexity of the $\mathbf{T_S}$ tuning mechanism we have not attempted to include it in the current model.

Finally, we would like to point out a similarity between the $LSNR_m$ formula presented here and one proposed by Van Doorn and Koenderink (1984). Their formula was empirically fitted to psychophysical data obtained by exploring the spatial and temporal summation (integration) of motion information using an apparent motion stimulus with $N_L >> 1$. They expected, but did not find, that the LSNR threshold would be a function of the product of target area and stimulus presentation time. The empirical formula that they discuss is

$$\text{(4.16)} \qquad SNR_{threshold} = \frac{1}{\frac{AT}{3000} - 1}$$

where **A** is the area of the stimulus and **T** is the stimulus duration. Our threshold measurement paradigm simultaneously manipulates the contrast of both the signal and obscuring noise patterns by changing the input LSNR (see equation 4.3). The populational response threshold level can therefore be written in terms of the input LSNR as

$$\text{(4.17)} \qquad LSNR = \frac{C^2_{signal}}{C^2_{noise}} = \frac{1}{g(\Psi, T_S, D_S, L, etc...)}$$

(where we assume that the externally added noise is much larger than the internal noise.) If we assume that populational response to stimulus duration is separable from other parameters, we can re-write equation (4.16) as

$$\text{(4.18)} \qquad LSNR = \frac{1}{i(T_S, D_S) * h(\Psi, L, etc...)}$$

where the function $i(\mathbf{T_S},\mathbf{D_S})$ represents the increase of populational response with $\mathbf{D_S}$ for any given value of $\mathbf{T_S}$. If we fix all stimulus parameters except $\mathbf{D_S}$, $\mathbf{C_{stimulus}}$ is represented by some minimum $\mathbf{D_S}$ value required to perceive the stimulus. If we call that $\mathbf{D_S}$ value $\mathbf{D_{S\text{-}min}}$ and combine equations (4.15), (4.15) and (4.18) we can write

$$\text{(4.19)} \qquad LSNR_m(D_S) = \frac{1}{\left(\frac{i(D_S)}{i(D_{S-min})}\right) - 1}$$

Equation (4.19) has a form very similar to the empirical formula in equation (4.16). The difference is that the improvement of response with time in our model is not linear. Finding an exact definition for $i(\mathbf{D_S},\mathbf{T_S})$ requires finding a solution to the differential equation given in the appendix, but an approximation good for small $\mathbf{T_S}$ values and wide current pulses would have a form something like

$$(4.20) \qquad i(T_s, D_s) = A(T_s) \cdot \left[1 - e^{\frac{-D_s}{\tau_{sat}}} \right]$$

where $A(T_s)$ represents the influence of contrast sensitivity and τ_{sat} is as discussed above. When τ_{sat} is large compared to D_s, the improvement of threshold with D_s is approximately linear with respect to stimulus duration so equations (4.18) and (4.19) provide similar predictions.

4.9 Conclusions

The temporal integration model presented here captures a number of properties seen in psychophysical measurements of the temporal integration of random dot apparent motion information. The model is based on the leaky integration of pulsatile motion detector responses to apparent motion stimuli, where the pulses correspond to the displacements in the stimulus. Temporal contrast sensitivity determines the shape of constant-stimulus-duration threshold curves for image exposure durations shorter than about 133 ms. The shape of the threshold curve for exposure durations longer than 133 ms is determined by the leaky integrator time constant and the shape of the current pulses emitted by the motion detector. The leaky integrator model exhibits threshold saturation behavior seen in the psychophysical data as well as the dependence on exposure duration of the time required for that saturation. Longer exposure durations result in a longer time-to-saturation because the leaky integrator discharges more between pulses. When the motion detector output pulses are far enough apart, there is no temporal integration and thus no threshold improvement. This corresponds to the lack of temporal integration in psychophysical threshold measurements using long exposure durations. Finally, the behavior of the psychophysical threshold curves across spatial displacement sizes indicates the influence of a response threshold mechanism which is an integral part of the temporal integration model.

Acknowledgments

R. E. Fredericksen was supported by a an International Brain Research Organization Research (IBRO) Fellowship. F. A. J. Verstraten was supported by the Biophysics Foundation of the Netherlands Organization for Scientific Research (NWO). We thank Marieta Braks for serving as an observer. Special thanks go to Peter Schiphorst and Alexandra Bellekom for designing and building the custom stimulus generation hardware, and to Rob van Weerden for taking care of the 1001 peripheral details required for use of the equipment during experimentation.

Bibliography

Adelson, E. H. and Bergen, J. R (1985) Spatiotemporal energy models for the perception of motion. *JOSA A*, **2**, 284-299.

Adler, B., Bock, O., and Grüsser, O.-J. (1981) Alpha-stripes: A spatial periodicity appearing in stroboscopically illuminated moving random dot patterns. *Vision Research*, **21**, 913-924.

Baker, C. L. Jr. and Braddick, O. J. (1985a) Eccentricity-dependent scaling of the limits for short-range apparent motion perception. *Vision Research*, **25**, 803-812.

Baker, C. L. Jr. and Braddick, O. J. (1985b) Temporal properties of the short-range process in apparent motion. *Perception*, **14**, 181-192.

Burr, D. C. (1981) Temporal summation of moving images by the human visual system, *Proceedings of the Royal Society of London*, **B 211**, 321-339.

Burr, D.C., Ross, J. and Morrone, M.C. (1986) Smooth and sampled motion, *Vision Research*, **26**, 643-652.

Emerson, R.C., Bergen, J.R. and Adelson, E.H. (1992) Directionally selective complex cells and the computation of motion energy in cat visual cortex. *Vision Research*, **32**, 203-218.

Fredericksen, R. E., Verstraten, F.A.J. and Van de Grind, W.A., (in press), Spatio-temporal characteristics of human motion perception, to appear in *Vision Research*.

Fredericksen, R. E., Verstraten, F.A.J. and Van de Grind, W.A., Temporal integration of random dot apparent motion in human central vision, submitted.

Kelly, D.H. and Burbeck, C.A. (1984) Critical problems in spatial vision, in CRC Critical Reviews in Biomedical Engineering, **10:2**, 125-177.

Koenderink, J. J. (1988) Scale time. Biological Cybernetics, 58, 159-162.

McKee, S. P. and Welch, L. (1985) Sequential recruitment in the discrimination of velocity. *JOSA A*, **2**, 243-251.

Mikami, A., Newsome, W.T., and Wurtz, R. H., (1986). Motion selectivity in macaque visual cortex. I. Spatiotemporal range of directional interactions in MT and V1. *Journal of Neurophysiology*, **55:6**, pp. 1308-1327.

Mikami, A., Newsome, W.T., and Wurtz, R. H., (1986). Motion selectivity in macaque visual cortex. II. Mechanisms of direction and speed selectivity in extrastriate area MT. *Journal of Neurophysiology*, **55:6**, pp. 1328-1339.

Mikami, A., Newsome, W.T., and Wurtz, R. H., (1986). Motion selectivity in macaque visual cortex. III. Psychophysics and physiology of apparent motion. *Journal of Neurophysiology*, **55:6**, pp. 1340-1351.

Morgan, M.J. (1979) Perception of continuity in stroboscopic motion: A temporal frequency analysis, *Vision Research*, **19**, 491-500.

Nakayama, K. and Silverman, G. H. (1984) Temporal and spatial characteristics of the upper displacement limit for motion in random dots. *Vision Research*, **24**, 293-299.

Reichardt, W. (1961) Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. In Sensory Communication, Rosenblith, ed. (Wiley, New York, 1961), pp. 303-317.

Robson, J.G., (1980) "Neural Images: The physiological basis of spatial vision," in Visual Coding and Adaptability, Ed. C.S. Harris, LEA, Hillsdale, New Jersey.

Scobey, R. P. and Kan, P. L. E. (1991) A horizontal stripe of displacement sensitivity in the human visual field. *Vision Research*, **31**, 99-109.

Snowden, R. J. and Braddick, O. J. (1989a) Extension of displacement limits in multiple-exposure sequences of apparent motion. *Vision Research*, **29**, 1777-1787.

Snowden, R. J. and Braddick, O. J. (1989b) The combination of motion signals over time. *Vision Research*, **29**, 1621-1630.

Snowden, R. J. and Braddick, O. J. (1990) Differences in the processing of short-range apparent motion at small and large displacements. *Vision Research*, **30**, 1211-1222.

Van Doorn, A.J., and Koenderink, J. J. (1982) Spatial properties of the visual detectability of moving spatial white noise. *Experimental Brain Research*, **45**, 189-198.

Van Doorn, A.J., and Koenderink, J. J. (1982) Temporal properties of the visual detectability of moving spatial white noise. *Experimental Brain Research*, **45**, 179-188.

Van Doorn, A.J., and Koenderink, J. J. (1984) Spatiotemporal integration in the detection of coherent motion. *Vision Research*, **24**, 47-53.

Van Doorn, A.J., and Koenderink, J. J. (1985) Perception of movement and correlation in stroboscopically presented noise patterns. *Perception*, **14**, 209-224.

Van de Grind, W. A., and Grüsser, O.-J. (1973) Temporal transfer properties of the afferent visual system, in Handbook of Sensory Physiology, Volume VII/3A, R. Jung, editor.

Van de Grind, W. A., Koenderink, J. J. and Van Doorn, A.J., (1986) The distribution of human motion detector properties in the monocular visual field. *Vision Research*, **26**, 797-810.

Van de Grind, W. A., Koenderink, J. J., Van Doorn, A.J., Milders, M.V., and Voerman, H. (accepted) Inhomogeneity and anisotropies for motion detection in the monocular visual field of human observers. To appear in *Vision Research*.

Appendix

The electric circuit simulations were performed using a software package called CAZm that was developed at the Microelectronics Center of North Carolina (MCNC) in Research Triangle Park, North Carolina. The circuit configuration does not require that specific circuit simulator; other simulators should provide the same results using other simulators. The resistor and capacitor values were the same in all model threshold curves shown and were set to the values of 5000 ohms and 0.00003 Farads. The convergence criteria for the simulations was always a relative tolerance of 10^{-9} and an absolute tolerance of 10^{-10}. Output voltage sequences were processed according to the equations given in the paper. The circuit output voltage is the solution to the differential equation

$$I_i(t) = \frac{V_o(t)}{R} + C \cdot \frac{dV_o(t)}{dt} \qquad \text{(A1)}$$

where

$$I_i(T_s, t) = \max_n \left[\text{pulse}(n, T_s, t) \right] \qquad \text{(A2)}$$

This is just the superposition of a series of shifted current pulses defined as

$$\text{pulse}(n, T_s, t) = u(t - nT_s) * e^{\frac{-\ln\left(\frac{t-nT_s}{\tau}\right)}{4\sigma}} \qquad \text{(A3)}$$

where u(t) is the unit step function. Note that the definition of $I_i(t)$ makes finding an analytical solution difficult.

The basic pulse shape is adapted from one proposed by Koenderink (1988). The original function was designed as a basis for filter functions that are causal in both time and scale-time, the latter being an extension of a similar concept, scale-space, that is used in the analysis of spatial images. The details of the extension are contained in Koenderink (1988) and the scale-space concepts can be found elsewhere (e.g. see Koenderink, 1984). As a filter kernel, the function is designed to gather information from the

past. This can be thought of in neuronal terms as projecting neuronal output into the future. The modification of the pulse shape used here consisted of a reversal of the time axis in the formula presented by Koenderink (1988). The pulse shape parameter values used for all simulations were τ = 120 ms and σ = 0.06.

Chapter 5*: Spatial summation and its interaction with the temporal integration mechanism

R. E. Fredericksen, F. A. J. Verstraten and W. A. van de Grind

Abstract

The combination of visual motion information over visual space was investigated using random dot apparent motion in which the random dots had a life time of a single displacement. The results show that spatial summation follows the form of probabilistic summation for a wide range of spatial pattern displacement sizes and that this form holds when changing either motion stimulus height or width. Probabilistic summation breaks down when the motion display region approaches a specific height and/or width. These values, termed the critical height and width, increase with spatial displacement size and can be interpreted as a measure of the dimensions of the motion detector's receptive field. The critical height is smaller than the critical width, a result that is consistent with a motion detector receptive field that is elongated in the direction of motion. Perhaps most importantly, the mechanisms of temporal integration and spatial summation are shown to work independently under the conditions tested here. Finally, the performance of the temporal integration mechanism appears to show evidence of an interaction between the form and motion perception systems.

5.1 Introduction

In primates, neurons in visual area V1 are sensitive to directional motion in the visual field. The directional information is then passed to other stages, perhaps most notably, area MT. We can consider directionally selective neurons, abstractly, as a large population of individual motion detectors, each of which calculates directional motion for some portion of the visual field. Each motion detector has some visual properties to which it responds best, including the direction of the motion and the scale of objects that move through that part of the visual field. We can think of more complex properties, like object motion, egomotion, or object shapes indicated by motion as resulting from further calculations based on the outputs of these individual motion detectors. The accuracy of these calculations can be affected to a great extent by noise in the detector signal. The output signal of each motion detector can include the influence of confounding motion noise in the visual field as well as noise injected into the signal internally (e.g. spontaneous neural firing). The output of a single detector is thus inherently ambiguous: we can not be absolutely sure whether the firing of the motion detecting neuron is due to motion in the visual field or just noise in the system. Nevertheless, our visual systems can reliably detect even very small visual motion signals. We know from signal processing theory that we can reduce the relative amount of noise by filtering the output signals of the detectors. Visual systems might do this both over visual space (spatial filtering) and over time (temporal filtering). The experiments presented here are designed to explore the biological mechanisms underlying these two types of filtering action.

5.1.1 Spatial Summation

Previous work on the spatio-temporal integration of motion information showed that, under some conditions, combination of motion information over space follows a form predicted by probabilistic summation (Lappin and Bell, 1976; van Doorn and Koenderink, 1984). For this reason we will refer to the spatial filtering of motion information as *spatial summation* throughout the rest of the paper. A probabilistic summation mechanism can be achieved through simple averaging (or adding) of detector outputs and is advantageous when the signal and noise components of the detector output are statistically independent. In this paper we are interested in how we can manipulate the visual motion stimulus in a way that will cause the spatial summation mechanism to reveal itself in the response of the detector population. We want a quantitative relationship

* The material in this chapter will be submitted to the journal *Vision Research* in almost identical form under the title "Spatial summation and temporal integration of visual motion" by the above authors.

between the number of detectors and the perception threshold. Two previous derivations of this relationship (Lappin and Bell, 1976; Van Doorn and Koenderink, 1982) were given in terms of the spatio-temporal correlation of the visual stimulus. The experimental paradigm used here measures perceptual thresholds by manipulating both the signal and the noise present in the motion stimulus. We can write the relationship between perceptual threshold (in terms of the input signal to noise ratio) and the number of independent detectors as

$$\frac{\text{Signal}_{in}^2}{\text{Noise}_{in}^2} = \frac{K}{\sqrt{N}} \tag{5.1}$$

In the experiments described below we use the quantity on the left as our threshold variable and look for the inverse square root relationship in the threshold results.

5.1.2 Temporal Integration

Visual motion experiments (both psychophysical and neurophysiological) are often performed using computer generated apparent motion stimuli. The term apparent motion refers to a visual motion percept induced by the presentation of a sequence of static images at a given rate. The directionally selective cells in primate cortex have been shown to respond to individual spatial displacements of luminance bars in their receptive fields and to have preferred spatial displacement sizes (Mikami, Newsome and Wurtz, 1986a,b). Apparent motion stimuli contain discrete spatial displacements of some spatial pattern and should therefore result in pulsatile responses in the population of directionally selective cells in cortex. In a previous paper (Fredericksen, Verstraten, and Van de Grind, submitted-b; chapter 4) we present evidence that the filtering of the population of motion signals over time follows the form of leaky integration of motion detector output. We show that the psychophysical threshold for directional perception of apparent motion stimuli can be modeled by a system in which the output signals of motion detectors that respond in this pulsatile manner are fed into leaky integrators. The leaky integrator output represents the perceptual signal, and the leaky integrator mechanism is the temporal filter for apparent motion stimulus. The leaky integrator output increases over time, but the output level depends on the frame rate (time between detector output pulses) and the total stimulus duration (the number of pulses). For this reason we will refer to the temporal filtering of motion information as *temporal integration* throughout the rest of the paper.

5.1.3 Spatial Summation and Temporal Integration in Concert

The temporal integration and spatial summation mechanisms can be related to known properties of neurophysiology and neuroanatomy. Neurons have been widely modeled as leaky integrators, and firing rates of directionally selective cells can increase over the duration of the stimulus (Mikami, Newsome and Wurtz, 1986a). The use of leaky integration as a temporal filter mechanism may therefore arise naturally at each cell in the visual pathway. Conversely, we know that visual motion is combined over space, under some conditions following a form indicated by equation (5.1). This could be accomplished through the normal pooling of neuronal responses resulting from convergence of neuronal outputs between visual processing stages (Van Doorn and Koenderink, 1984). It would be difficult or impossible to try to locate these properties at individual neuronal or anatomical sites because the system is implemented in a distributed manner. Nevertheless we can discuss the properties as arising within the motion detector system.

In figure 5.1 we have shown the spatial summation step as a uniformly weighted summation of detector output signals. The summed signal is then input to a leaky integrator stage. We note that the linear nature of this first order model allows us to reverse the order of the mechanisms. However, neither order is more advantageous because the two mechanisms are likely to be intertwined throughout the physical system and are sure to have non-linear properties. The model in figure 5.1 is therefore a simple guide for our thinking. The point most important to the current results, however, is that the differing origins of the two mechanisms indicates that they should operate independently.

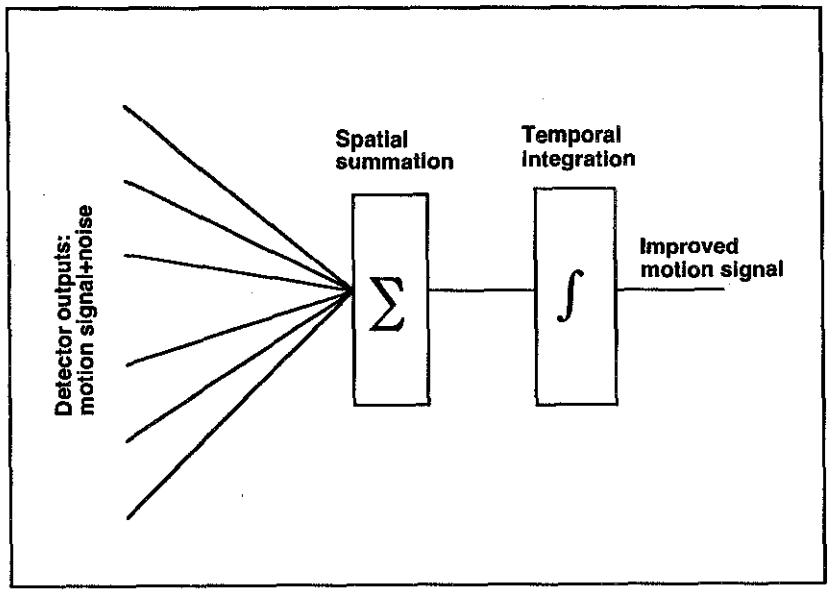


Figure 5.1
The proposed model for the interaction of spatial summation and temporal integration. The spatial summation step is shown as a uniformly weighted summation of detector output signals. The temporal integration step is shown as a leaky integration stage.

5.1.4 Experiment Design

We will use a model of the layout of the motion detector receptive fields shown in figure 5.2. The model has distinct detector populations tuned to a single direction of motion (for modeling convenience) but with different preferred spatial displacement sizes. The receptive fields of biological detectors would overlap and would not be so regular in their coverage of the visual field, but figure 5.2 is a convenient model. The receptive field size is shown as increasing with preferred spatial displacement size, and each population covers the same portion of the visual field.

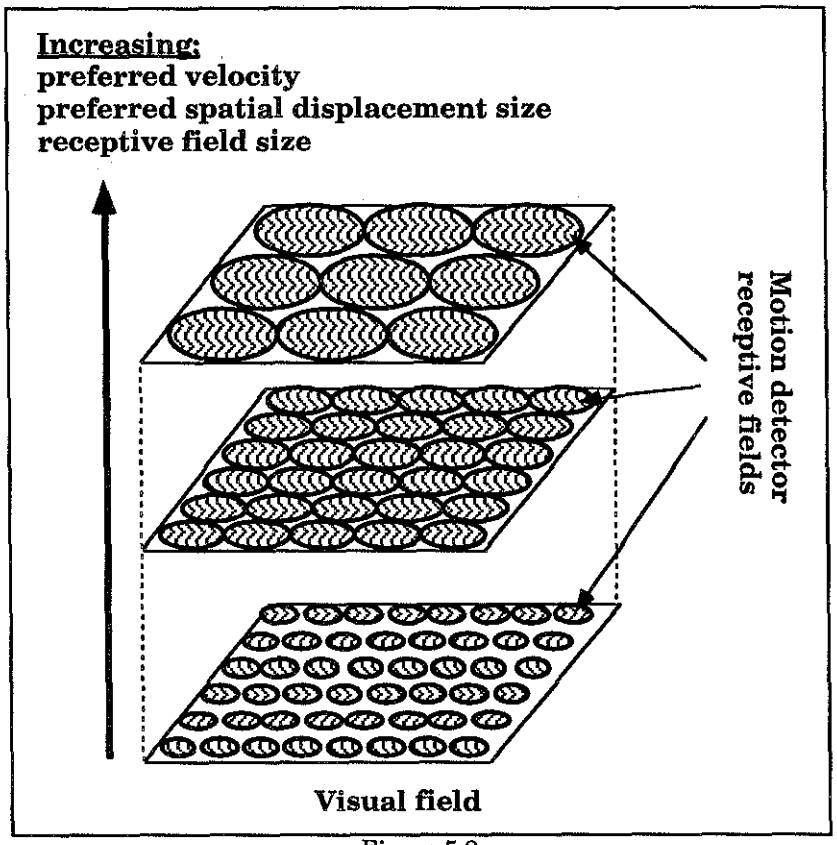


Figure 5.2
The model of the receptive field layout for motion detectors serving one region of the visual field. Receptive fields are shown as increasing in size with preferred spatial displacement of the spatial pattern, and with increasing velocity. The receptive fields are shown segregated by size. Only three sizes are shown, but the biological system is likely to have a continuous distribution of receptive field sizes.

Changing the area of the stimulus changes the number of detectors reached by the stimulus (see figure 5.2) and is equivalent to changing the number of detectors that feed the spatial summation and temporal integration mechanisms in figure 5.1. Van Doorn and Koenderink (1984) show that reducing the stimulus area changes the LSNR threshold. In their stimulus, when the area of the display that contained coherent motion was reduced, the removed areas were filled in with incoherent

motion. Thus the total display area was constant, but the height and width of the region of the stimulus containing coherent motion was changed. When we refer to changing stimulus height or width throughout the rest of the paper, we are referring to a change in the region of the stimulus containing coherent motion, and not the overall stimulus display.

Van Doorn and Koenderink (1984) found that, for low velocity (small S_S) random dot patterns moving horizontally, changing the height of a square stimulus resulted in a variation of LSNR threshold that follows the form predicted by equation (5.1). Changing the width of the stimulus produced somewhat different results for all velocities used. They postulated that the changing width experiment not only reduces the number of detectors responding to the stimulus, but changes the population of detectors responding to the stimulus. Their stimulus consisted of a random dot pattern that was repeatedly displaced across the screen. That stimulus includes spatio-temporal correlation between image frames that are not adjacent in the image sequence, thereby providing spatial displacement information at integer multiples of the basic spatial displacement size of the stimulus. This means that multiple subpopulations of detectors in figure 5.2 were responding to the stimulus. Detectors with larger preferred spatial displacement sizes also have larger receptive fields (Van de Grind, Koenderink and Van Doorn, 1986; Mikami Newsome and Wurtz, 1986b) that appear to be elongated in the direction of motion (Van de Grind, Koenderink and Van Doorn, 1986). This means that as the width of the display is changed, the number of detectors responding to that stimulus is reduced by removing entire classes of detectors all at once. This becomes clearer if we consider what happens when the width of the motion display approaches the width of the receptive field of the detector subpopulation. The width of the stimulus must be the same as or just less than the width of the individual detector for that detector to respond. Once the detector is narrow enough, none of the detectors of that receptive field size (or larger) will respond to the stimulus. Thus as the screen size is reduced, detectors tuned to the larger displacement sizes in the stimulus are removed as a group (because of the discrete nature of the stimulus) rather than in a linear fashion.

In a previous paper we employ a stimulus paradigm that isolates detector populations by preferred spatial displacement size and frame rate (Fredericksen, Verstraten and Van de Grind, in press; chapter 2). This is equivalent to selective stimulation of only one of the subpopulations shown in figure 5.2. Changing the width of the display for such a stimulus should therefore have similar results as changing the height of the display, up to the point where the stimulus width approaches the detector receptive field width. We therefore use the same experimental approach as Van Doorn and Koenderink but use the new stimulus paradigm. The results using the new stimulus indicate that equation (5.1) holds for both changing height and changing width as predicted by the model in figure 5.2. The data further confirm previous evidence that the receptive field size of the motion detectors increases with the preferred spatial displacement size and that the receptive fields are elongated in the preferred direction of the detector. Tests of the interaction of spatial summation and temporal integration show that the two mechanisms can indeed operate independently. Finally, there appears to be evidence for a short term inhibitory effect from the high contrast edges of the useful display area that affects the visibility of the motion.

5.2 Methods

5.2.1 Stimulus Generation

The motion stimuli were generated using custom image generation hardware driven by a Macintosh IIfx computer. The monitor had a P4 phosphor and a base frame rate of 90 Hz. All stimulus frame rates were integer multiples of the base frame rate. The interstimulus interval was negligible in all experiments. The display screen was 14 centimeters and 256 pixels square, and pixels were 0.55 mm in diameter. Experiments were performed with the display at a distance of either 1.0 or 4.0 meters from the subject. The useful display region was 2.0 degrees square at 4 meters and 8.0 degrees square at 1 meter. A black fixation dot (approximately 2 minutes of arc in diameter) was affixed to the center of the stimulus display area. All experiments employed a luminance signal-to-noise ratio (LSNR) threshold method. Threshold

measurements were produced by modulating the luminance of the target random dot pattern through addition of a spatially and temporally uncorrelated noise random dot pattern. Some reasons for use of such a paradigm have been enumerated in Fredericksen et al. (in press; chapter 2) and the method has been described in detail elsewhere (van Doorn and Koenderink, 1982a, 1982b). The average luminance of the signal-plus-noise random dot patterns was set to 50 cd-m^{-2} in all experiments. The signal-plus-noise random dot pattern was set to an average contrast level of 70%, a value well above the motion threshold saturation value of 35% (van de Grind, Koenderink and van Doorn, 1987).

5.2.2 Threshold Measurement

Directional motion thresholds were measured using a two-alternative-forced-choice paradigm (motion to the left vs. motion to the right) to distinguish a direction discrimination threshold from a simple motion, contrast change, or flicker detection threshold. Left-right motion was used to discount anisotropies in motion direction discrimination ability (Scobey and Kan, 1991; van de Grind, Koenderink, van Doorn, Milders and Voerman, in press). In all experiments the stimulus presentation consisted of a number of frames of random dot pattern motion superimposed on a mean luminance background. The thresholds were determined using a staircase procedure that tracked the 79% correct level; three consecutive correct directional discriminations resulted in a lowering of the LSNR value, while any other sequence raised the LSNR value by the same amount. There were 10 turning points in all experiments with the threshold calculated as the average of the final six turning points. A step size of 2 dB was used for the first 4 turning points and a step size of 1 dB was used for the final six. Turning point sequences were inspected and any sequence that did not reach a stable fixed point due to attentional variation was rejected and re-measured. In order to decrease the required experimental time, the subjects were asked to manipulate the threshold variable manually (before beginning each staircase procedure) until it was just above subjective threshold. If a subject mis-estimated the starting value, that mis-estimation was apparent in the turning point sequence of the staircase procedure as a continuing upward trend in the turning points. Such sequences were rejected and the point re-measured. Subjects were allowed to rest after any or all of the individual threshold measurements. The experiments were performed in a portion of a room screened from stray light with ambient lighting provided by the display screen. All experiments were monocular, and the subjects used a chin and head rest. Subjects were instructed to fixate on the black dot on the screen and to determine either the direction of motion or the side of the display on which the stimulus was presented. The subject's answer was recorded via a key press on the computer keyboard and the threshold variable was automatically changed by the computer. All subjects had normal or corrected to normal vision.

5.3 Results

We begin with a short description of the stimulus and a summary of symbols that will be used throughout the rest of the paper. The stimulus is composed of a sequence of spatially displaced random dot patterns displayed on a video display screen. The stimulus was designed with a functional model of human motion perception as a guide that is essentially reproduced in figures 5.1 and 5.2, with the additional use of bi-local detectors as the actual motion detector mechanism (Fredericksen, Verstraten and Van de Grind, in press; chapter 2). To simplify discussion of the stimulus, we define symbols for some stimulus parameters. The frame duration (spatial pattern exposure time) and spatial pattern displacement size in the stimulus are referred to as $\mathbf{T_S}$ and $\mathbf{S_S}$, respectively. *Frame rate* is the inverse of $\mathbf{T_S}$, and the two terms will be used interchangeably throughout the paper. Displacement count ($\mathbf{N_D}$) is the number of spatial displacements in the stimulus, or one less than the number of frames in the image sequence. The stimulus duration ($\mathbf{D_S}$) in our apparent motion stimulus is defined as frame duration ($\mathbf{T_S}$) times the number of frames ($\mathbf{N_D}$ + 1) because we do not use an interstimulus interval (presentation of a uniform spatial luminance pattern between the spatially displaced pattern images). The dot life time $\mathbf{N_L}$, the number of coherent spatial displacements that a dot in the spatial pattern undergoes before it is randomly displaced, controls the spatio-temporal displacement content of the stimulus. We used $\mathbf{N_L}$ = 1 in all directional motion detection experiments because it limits the coherent

motion information contained in the stimulus to a single $<S_S,T_S>$ combination. It is control of N_L that allows selective stimulation of the detector subpopulations (Fredericksen, Verstraten and Van de Grind, in press; chapter 2). Finally, our directional motion detection experiments use a luminance signal-to-noise ratio method to determine the perceptual threshold. We will refer to the results as LSNR thresholds, or just threshold curves.

5.3.1 Experiment 1

The first set of experiments were designed to further explore the probabilistic summation mechanism previously proposed (Van Doorn and Koenderink, 1984). We use the same experimental approach: measure threshold values with either motion region height or width as the independent variable. The display layout is shown in figure 5.3. The useful display area was square and surrounded by a dark, non-display region of the monitor. The luminance of the dark area of the monitor was due only to reflected ambient light. The useful display area was composed of zero contrast, mean luminance regions flanking a 70% average contrast motion region. The white regions in the figure represent the mean luminance regions and the dotted-texture portion represents the motion display region. The figure shows a configuration for a changing height experiment. A changing width experiment would be represented by a vertically oriented strip of dotted texture. The height or width of the motion region was changed by incrementally replacing the edges of the motion region with zero contrast, mean luminance regions, thereby reducing the height or width of the motion region. In the first experiment the display screen was at a distance of 1 meter from the observer. The useful display area was therefore 8 degrees square. The stimulus duration was 1000 ms for subject FV and 700 ms for subject EF. (A value of 1000 ms was used for subject FV so that the threshold values could be compared in experiment 3.) Both subjects were tested at three values of S_S that are within the visible range for both subjects (e.g. see Fredericksen, Verstraten and Van de Grind, in press; chapter 2) with T_S set to 33 ms.

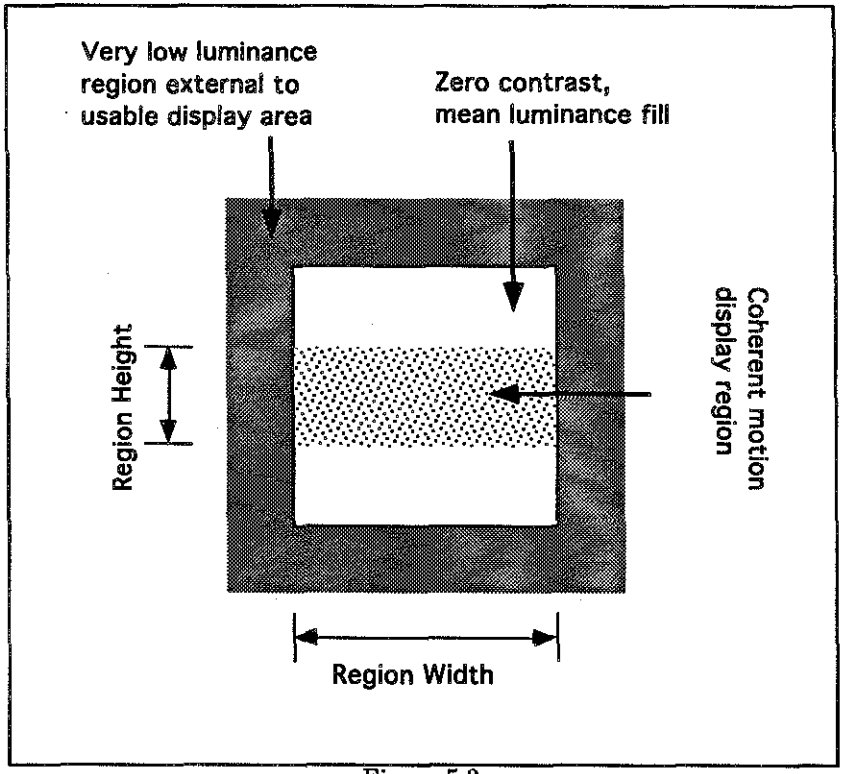


Figure 5.3
Experiment 1 display configuration. The useful display area was square and surrounded by a dark, non-display region of the monitor. The luminance of the dark area of the monitor was due only to reflected ambient light. The useful display area was composed of zero contrast, mean luminance regions flanking a 70% average contrast motion region. The white regions in the figure represent the mean luminance regions, and the dotted texture portion represents the motion display region. The figure shows a configuration for a changing height experiment. A changing width experiment would be represented by a vertically oriented strip of dotted texture. The height or width of the motion region was changed by incrementally replacing the edges of the motion region with zero contrast, mean luminance regions, thereby reducing the height or width of the motion region.

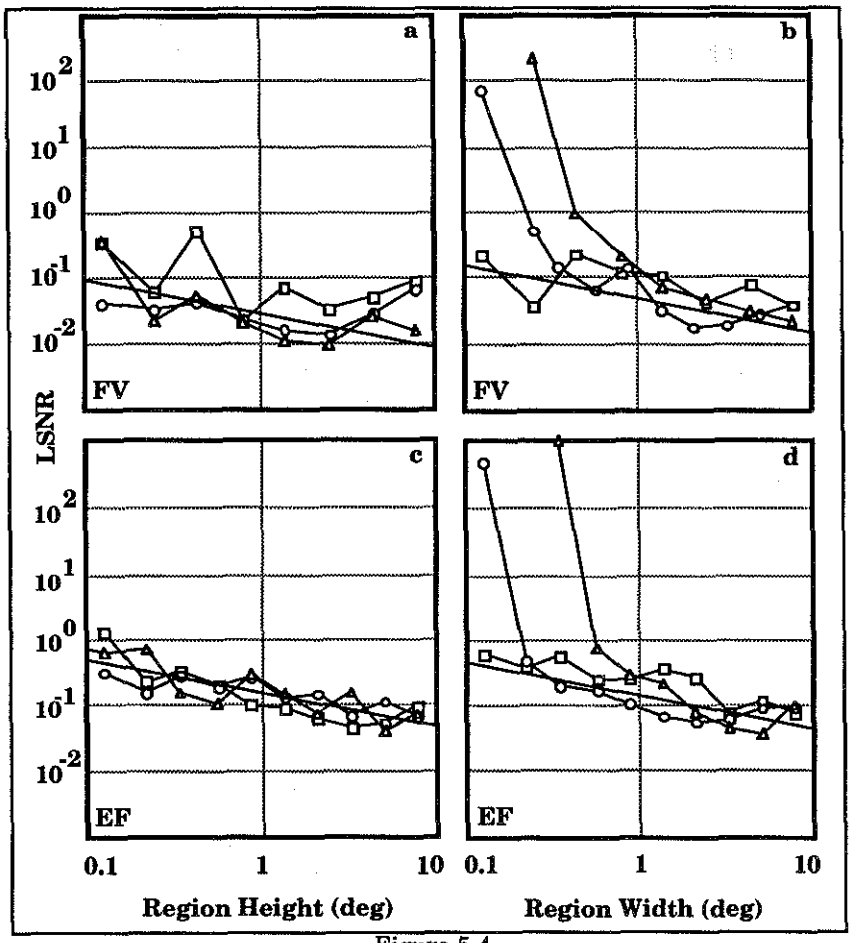


Figure 5.4
Experiment 1 results. The vertical axis is LSNR threshold value while the horizontal axis is the width or height of the motion display region as labeled in each graph. The other dimension of the stimulus was held constant at full size. The symbols in the graphs indicate the use of different S_S values: □ = 0.031 degrees (1.9 minutes), ○ = 0.094 degrees (5.6 minutes), and △ = 0.188 degrees (11.3 minutes). Figure 5.4a and 5.4b show the threshold curves for subject FV for varying motion region height and width, respectively. Figures 5.4c and 5.4d show the same data for subject EF.

The data are shown in figure 5.4. The vertical axis is LSNR threshold value, while the horizontal axis is the width or height of the motion display region as labeled in each graph. The other dimension of the stimulus was held constant at full size. Figure 5.4a and 5.4b show the threshold curves for subject FV for varying motion region height and width, respectively. Figures 5.4c and 5.4d show the same data for subject EF. The relationship in equation (5.1) can be represented on a log-log plot of LSNR threshold versus detector count (stimulus height or width) as a line of slope -0.5. The threshold data are therefore displayed on a log-log scale with a line of that slope added to the graph as a reference.

5.3.2 Experiment 1 Discussion

It is clear that the threshold value depends differently on the height and width of the stimulus. Likewise, the height and width curves behave differently depending on the value of S_S. The height threshold curves generally follow the theoretical line for all of the display heights tested here. The width threshold curves, however, eventually move away from the theoretical curve, decreasing in visibility at a faster rate than predicted by probabilistic summation (equation 5.1). This represents a rate of loss of visibility of the stimulus that is not accounted for by. Following similar usage of the term by Van Doorn and Koenderink, we refer to the region width at which this additional loss begins as the critical width. The critical width increases with increasing S_S. This may correspond to the increase in receptive field width with preferred displacement size of the detector. We note here that both a critical width *and* a critical height appear in the threshold curves presented by Van Doorn and Koenderink (1984) (cf. their figures 3 and 4).

Finally, both the height and width curves seem to show a paradoxical decrease in threshold (increase in visibility) of the stimulus as the stimulus size is initially decreased from its maximum dimensions. This was not a motion adaptation effect because varying the sequence of stimulus sizes during threshold measurement did not remove it. An alternative hypothesis is that it is the result of some kind of influence from the edge of the useful display area. The surround of the 8 degree display area was of very low luminance, only reflecting ambient light, while the useful display region of the screen was an average of 50 cd-m^{-2}.

Thus there was a high contrast edge surrounding the display area. When the height or width of the motion region size was reduced, the motion region moved farther from this high contrast edge. The data presented by Van Doorn and Koenderink did not show this kind of behavior. In those experiments the largest motion region dimension was (usually) 2 degrees while the incoherent motion surround (equivalent to our zero contrast surround) was 5 degrees square. We postulate that this practice reduced or eliminated the edge effects seen in the current data.

5.3.3 Experiment 2

The next experiments were designed to test the possible influence of the high-contrast edge of the display and to explore the lack of a critical height in the current results. Subject FV tested the contrast edge influence hypothesis by repeating experiment 1 for a much smaller motion display area as shown in figure 5.5. The display was exactly the same as in the first experiment except that the motion region was restricted to the central 2 degree square of the 8x8 degree display region. The motion region was therefore always far from the high contrast edge of the useful display area. Subject EF repeated experiment 1 at a distance of 4 meters, thereby reducing the useful display area to 2 degrees square. The other stimulus parameters were the same as in experiment 1 for both subjects.

5.3.4 Experiment 2 Discussion

Figures 5.6a and 5.6b show the height and width threshold curves for subject FV. Figures 5.6c and 5.6d show the height and width threshold curves for subject EF. The data is graphed in the same manner as figure 5.4. It appears that in both cases the paradoxical loss of visibility for the largest display sizes is reduced but not completely removed. We expected subject FV's data to show no more edge effect, and subject EF's data to show a clear but somewhat smaller edge effect because subject FV used a display that included a large surround area of mean luminance.

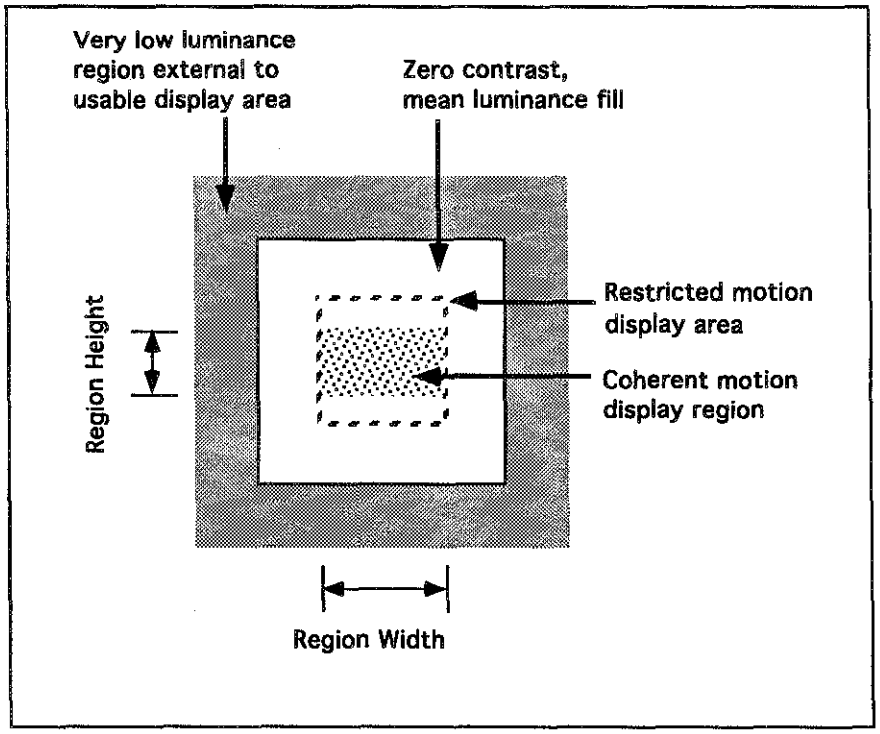


Figure 5.5
Experiment 2 display configuration. The display was exactly the same as in the first experiment for subject EF. It was the same for subject FV, except that the motion region was restricted to the central 2 degree square of the 8x8 degree display region.

At first glance, it seems that the critical width values have increased for subject FV but decreased for subject EF. However we point out that the sampling of the region dimensions has changed and that the changes in critical width are within a range of error that could result from that change (aliasing effects). It is interesting to note that subject EF now shows a critical height for the largest S_s value. Subject FV also appears to show some evidence of a critical height for the largest S_s value. Finally, we note that the 2 degree square configuration results in a motion display area that is only 25% of the motion display area for the same height/width values in the 8 degree square configuration. The resulting difference in the threshold values for each individual is

consistent with equation (5.1). The four-fold difference in motion region size appears as an average 0.3 log-unit translation[1] of the threshold curves up the vertical axis.

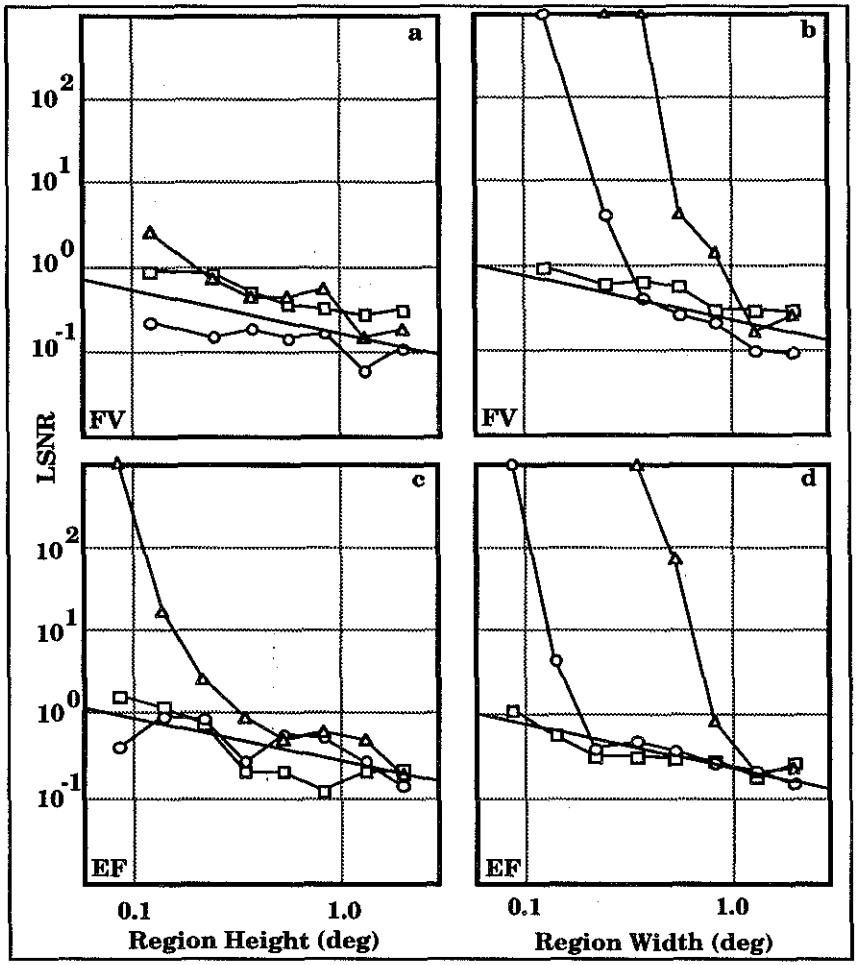


Figure 5.6
Experiment two results plotted as in figure 5.4.

[1] The LSNR improves inversely with the square root of the area, and we are plotting the log of the LSNR: $\log\left(\sqrt{4}\right) = 0.301$.

5.3.5 Experiment 3

The final experiment was designed to look for interactions between the temporal integration and spatial summation mechanisms. This was achieved by repeating the first experiment using additional $\mathbf{D_S}$ values of 100, 200, 400, 600 ms. All other stimulus parameters were the same as in experiment 1.

5.3.6 Experiment 3 Discussion

Figure 5.7 shows the results for subject FV plotted along with the $\mathbf{D_S}$ = 1000 ms curve from experiment 1. The left column shows the height curves, and the right column shows the width curves. The $\mathbf{S_S}$ value is increasing down the columns. The parameter of the curves in each graph is $\mathbf{D_S}$. The probabilistic summation behavior of the curves is most prominent for the $\mathbf{D_S}$ = 1000 ms curves. As $\mathbf{D_S}$ increases from 100 to 1000 ms, the threshold levels move toward values that are more consistent with probabilistic summation. The paradoxical effect is characterized by a loss of visibility at the largest display sizes. The effect is larger for small $\mathbf{S_S}$ than for large $\mathbf{S_S}$ (the top two graphs in figure 5.7), and for short stimulus durations. As $\mathbf{D_S}$ increases, or alternatively as the threshold saturates (stops decreasing), the effect also decreases. It is important to note that if the mechanism behind the effect was a constant reduction of visibility (inhibition of the response of the detector population), the final LSNR saturation level would be much closer to the shape of the shortest $\mathbf{D_S}$ threshold curves. Instead, it seems that the paradoxical effect is removed or reduced as duration increases, although the effect still appears in some of the 1000 ms curves (figures 5.7a and 5.7c).

The paradoxical effect does not significantly change the adherence of the 1000 ms threshold curve to equation (5.1) except perhaps for large display sizes. The width curve for the smallest $\mathbf{S_S}$ value at 100 ms (figure 5.7b) has a very strong positive slope whereas the 1000 ms curve follows very closely the predicted -0.5 slope. As $\mathbf{D_S}$ is increased, the curves become generally flat for display sizes larger than the critical dimensions. This flat region may represent either a balancing of the paradoxical effect with that of spatial summation, or a combination of parameters in which spatial summation is not in effect. In either case, there is a significant

change of the threshold levels that depends on both stimulus duration and display size.

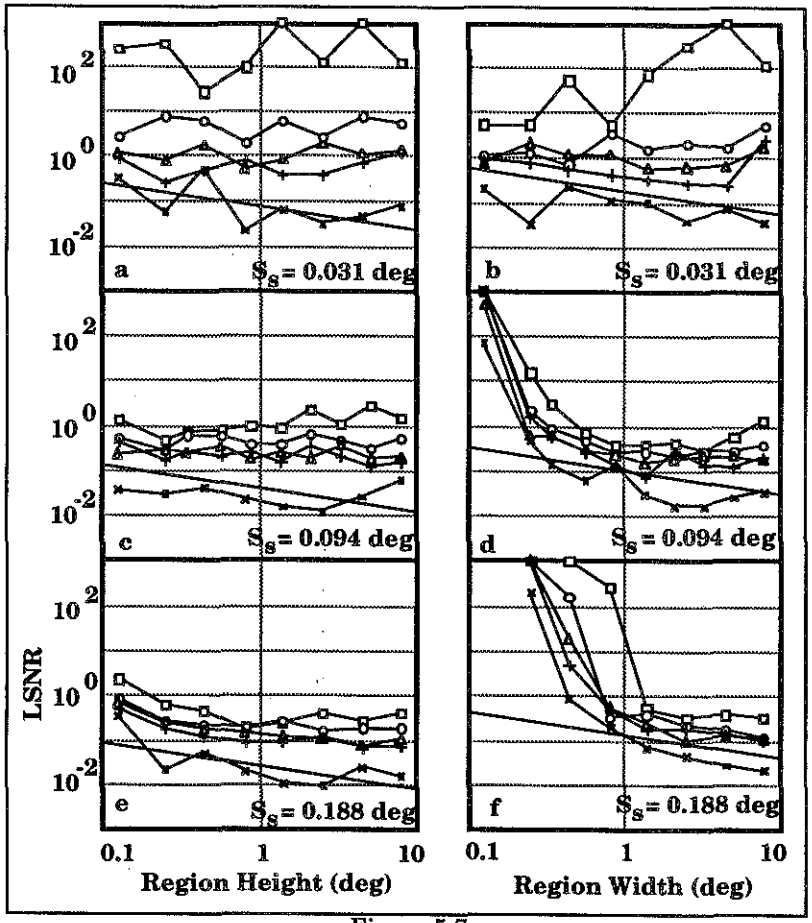


Figure 5.7

Experiment 3 results for subject FV plotted along with the D_S = 1000 ms curve from experiment 1. The left column shows the height curves, and the right column shows the width curves. The S_S value is increasing down the columns. The parameter of the curves in each graph is D_S. The symbols in the graphs indicate the use of different D_S values: □ = 100 ms, ○ = 200 ms, Δ = 400 ms, + = 600 ms, and x = 1000 ms.

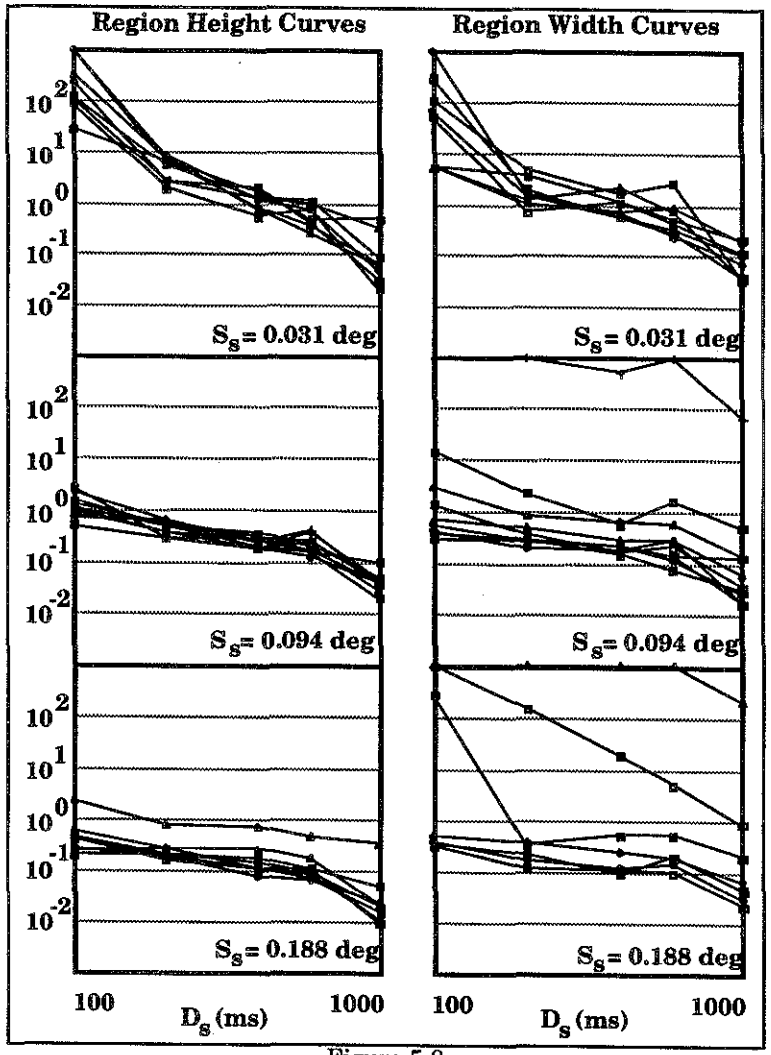


Figure 5.8

The data in figure 5.7 are re-plotted as a function of D_S for constant stimulus dimensions. The S_S value is increasing down the columns. The parameter of the curves in each graph is stimulus height or width. We do not identify different heights and widths within each graph because it is clear that they behave identically except when the stimulus height or width causes the motion to be non-perceivable.

Figure 5.8 shows the data in figure 5.7 re-plotted as a function of D_S. The left column shows the height curves and the right column shows the width curves. The S_S value is increasing down the columns. The parameter of the curves in each graph is the relevant stimulus dimension of height or width. We do not identify different heights and widths within each graph because it is clear that they behave identically except when the stimulus height or width causes the motion to be non-perceivable. The data shows that the temporal integration time constant does not depend on region height or width, and thus the number of stimulated detectors. Furthermore, the curves are consistent with those in our previous experiments exploring temporal integration (Fredericksen, et al., submitted-a,b; chapters 3 and 4).

5.4 General Discussion

We first discuss the relationship of the current results with those of Van Doorn and Koenderink (1984). Both experiments show the same critical width and critical height behavior in the threshold curves. Note that the higher velocities used by Van Doorn and Koenderink were created through the use of large spatial displacement sizes at the basic frame rate of 100 Hz. When S_S is greater than one half of the largest visible S_S for a subject, there is no perceivable coherent spatio-temporal correlation between frames that are not adjacent in the image sequence. Under such conditions the multiple displacement stimulus is very much like the N_L = 1 motion stimulus used here. This explains why their high velocity and our large displacement size curves look very similar.

5.4.1 Critical dimensions of the stimulus

Our results and those of Van Doorn and Koenderink show the same critical width and critical height dependence on spatial displacement size. Larger S_S values have larger critical widths, and the largest S_S values have a critical height. Note that both subjects in the current experiments show approximately the same critical height and width values and that the critical height is less than critical width. In the lowest velocities in their data and the smallest S_S values in our data the critical dimensions do not appear . Van Doorn and Koenderink postulate that the critical dimensions correspond to the basic receptive field dimensions for the motion detectors tuned to that velocity. The critical height is considered an indication of the size of the receptive field subcomponents of the bi-local detector, while the critical width is an indication of the separation of those subcomponents in the visual field. Such an explanation is consistent with the previous and current results, especially if spatial summation within the basic receptive field of the motion detector follows some other form than probabilistic summation. However, there is an alternative explanation that must be considered. When the spatial pattern is displaced, we must fill in one side of the display with fresh random dots. This causes a reduction in the number of correlated dots during the spatial displacement. The ratio of correlated to uncorrelated dots in our stimulus, which we will call the spatial signal-to-noise ratio (SSNR), is

$$\text{SSNR} = \frac{W - S_S}{W + S_S} \tag{5.2}$$

where W is the width of the motion region. Note that this quantity is constant for the region height curves because the region width is fixed in those experiments. However this value changes across the region width curves. The same effect applies to the stimuli used by Koenderink and Van Doorn (1984). The SSNR for the stimuli that they used ($N_L \gg 1$, $N_D \gg 1$) can be calculated as

$$\text{SSNR} = \frac{W - S_S}{S_S} \tag{5.3}$$

For a given S_S, fixing the region width fixes the SSNR. The LSNR and SSNR measures are somewhat different, and it is not clear how they interact to determine the final perceptual threshold.

In a different approach, Lappin and Bell (1976) measured the accuracy of directional motion detecting using two frame apparent motion ($N_D = N_L = 1$). The formula that they derive for the total spatio-temporal correlation is

$$\text{TotalCorrelation} = A \cdot \left(\frac{W - S_s}{W}\right) \cdot \sqrt{\text{StimulusArea}} \tag{5.4}$$

where **A** is a constant. Plotting the LSNR thresholds in figure 5.7b and 7d against the value of TotalCorrelation does not adjust the curves enough to account for the critical width value (graphs not shown), so the influence of the SSNR is not a simple linear factor. Plotting the LSNR thresholds as a function of equation (5.2) (graphs not shown) indicates that, for stimulus widths near the critical width, the influence is at least inversely proportional to the fourth power of the SSNR. We can therefore hypothesize an extension of equation (5.1) for changing motion region width that includes the influence of SSNR to be

$$\frac{\text{Signal}_{in}^2}{\text{Noise}_{in}^2} = \frac{K}{\text{SSNR}^4 \cdot \sqrt{N}} \tag{5.5}$$

Lappin and Bell (1976) derive their equation from first principles while the influence of SSNR in equation (5.5) is an empirical estimate. However the two may not actually be at odds. It is clear from equation (5.2) that the SSNR decreases as **W** approaches $\mathbf{S_S}$. However, it is not clear whether the critical width behavior is a result of increasing SSNR or of a mismatch between the stimulus and motion detector receptive field dimensions. Our empirical estimate is presented as a function of total spatio-temporal correlation in the stimulus, but its form does not follow that predicted by theory. The sources of spatial noise and luminance noise are independent, so we would expect them to interact in a much less drastic manner, assuming that they both influence perceptual threshold by reducing the spatio-temporal correlation in the stimulus (reducing the overall SNR of the stimulus). Furthermore, the total SSNR stays constant for the height curves, so a changing SSNR can not account for the critical height behavior of the threshold curves. We therefore believe that the critical dimension behavior must at least in part be due to the similarity of the stimulus dimensions to those of the receptive field of the basic motion detectors as suggested by Van Doorn and Koenderink (1984). The logical consequence of this conclusion is that the motion detector receptive field size (and hence the sizes of that motion detectors front-end receptive fields) increases with increasing preferred spatial displacement size (or $\mathbf{S_S}$), and that the motion detector receptive fields are elongated in the direction of motion.

5.4.2 Velocity dependence of the critical dimensions

It appears that the basic receptive field dimensions follow the size of $\mathbf{S_S}$ in the motion stimulus rather than velocity. If the critical dimensions are proportional to velocity, then decreasing velocity should decrease the critical dimensions. To test this hypothesis, subject EF repeated experiment 2 with $\mathbf{T_S}$ = 66 ms, reducing the three equivalent velocities ($\mathbf{S_S/T_S}$) by a factor of two. The velocity hypothesis predicts that the largest $\mathbf{S_S}$ value in the experiment, being twice that of the middle $\mathbf{S_S}$ value, should show critical dimensions comparable to those for the middle $\mathbf{S_S}$ value in experiment 2. The data are shown in figure 5.9. The only difference from the $\mathbf{T_S}$ = 33 ms threshold curves is that the critical height for the largest $\mathbf{S_S}$ value may be somewhat *larger* than when $\mathbf{T_S}$ = 33 ms (again, we must be aware of possible region height sampling artifacts).

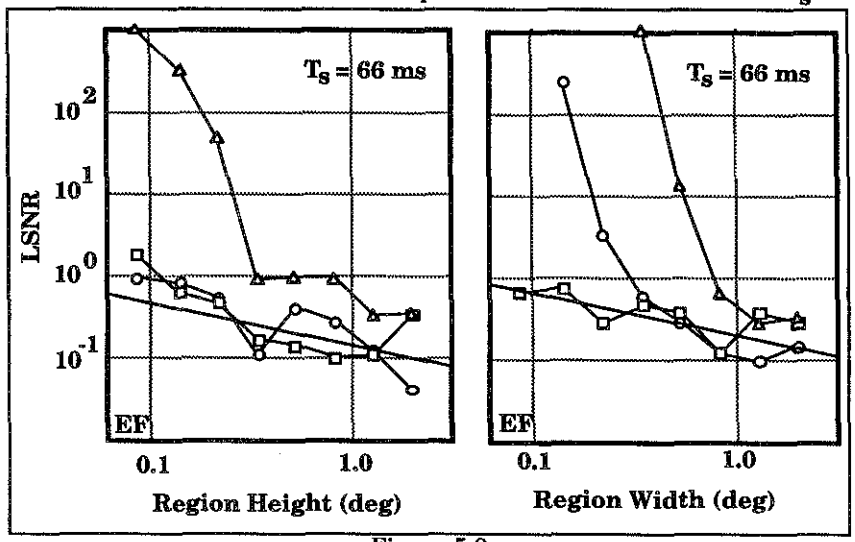


Figure 5.9
The results of the velocity hypothesis test. $\mathbf{T_S}$ was set to 66 ms rather than 33 ms, thereby reducing all stimulus velocities by a factor of two. The data are plotted as in experiment 2 for subject EF.

This result is counter to the velocity hypothesis. However the possible change is consistent with previous evidence that EF is more sensitive to large displacement values when T_S = 33 ms than when T_S = 66 ms (Fredericksen et al., in press; chapter 2).

5.4.3 Independence of the mechanisms

Van Doorn and Koenderink expected a linear relationship between stimulus duration and threshold. However our previous work (Fredericksen et al., submitted-b; chapter 4) indicates that the temporal integration mechanism for apparent motion corresponds to leaky integration of pulsatile motion detector outputs, and the current results follow the same qualitative and quantitative behavior. The current results also indicate that the spatial summation and temporal integration mechanisms work independently from one another, at least for the range of stimulus parameters used in the experiments presented here. The independence is especially clear when LSNR is low: a condition in which the target motion pattern is of very low contrast. These conditions are consistent with the linear range of the contrast response function of cat lateral geniculate cells (see discussion above). The independence of the two mechanisms should allow us to combine models of the two mechanisms (see [Fredericksen et al., submitted-b; chapter 4] for a model of temporal integration).

5.4.4 Paradoxical effect

We have already noted that if the cause was a simple, constant reduction of visibility due to the high contrast edge at the sides of the useful display, then the final LSNR saturation level would show the effects. Instead the effect decreases with stimulus duration, with the distance of the motion display region from the high contrast edge, and with the size of the high contrast region of the display. There are many possible explanations which can not be verified or falsified by the data presented here. For example, the effect might represent a competitive interaction of motion detectors that is reduced as stimulus duration increases. A second possibility is that the motion detectors at the edge of the display are affected by the high contrast edge, and may signal either slow or no motion. The experiments presented by Van Doorn and Koenderink used random noise fill rather than a flat luminance (as used here). A third possibility, then, is that the effect is due in part to the motion vs. no-motion edge at the intersection of the flat luminance region and the motion display region (C. A. Burbeck, personal communication, April, 1993). Understanding of this phenomenon will require investigations of a type beyond the scope of the experiments presented here.

5.5 Conclusions

Our experimental results confirm a number of conclusions reached by Van Doorn and Koenderink (1984) but also extend them. The combination of motion detectors over visual space follows very closely the form of probabilistic summation for a wide range of spatial pattern displacement sizes (stimulus velocities). Probabilistic summation holds for both changing height, as shown previously (Van Doorn and Koenderink, 1984), but also holds when the width of the stimulus is changed. We attribute this difference to the use of a single displacement lifetime dot stimulus. The breakdown of probabilistic summation reported by Van Doorn and Koenderink (1984) also occurs here. Motion visibility is lost when the motion display region approaches a height and/or width that depends on spatial displacement size (the critical dimensions). The data presented here are also consistent with the previous research in that the critical dimensions increase with spatial displacement size and are consistent with motion detector receptive fields that are elongated in the direction of motion. Finally, we show that the mechanisms of temporal integration and spatial summation are shown to work independently under the conditions tested here. Again, this may be a result of our single displacement lifetime stimulus paradigm.

Acknowledgments

R. E. Fredericksen was supported by a an International Brain Research Organization Research (IBRO) Fellowship. F. A. J. Verstraten was supported by the Biophysics Foundation of the Netherlands Organization for Scientific Research (NWO). We thank Marieta Braks and Richard van Wezel for serving as observers. Special thanks go to Peter Schiphorst and Alexandra Bellekom for designing and building the custom stimulus generation hardware, and to Rob van Weerden for taking care of the 1001

peripheral details required for use of the equipment during experimentation.

Bibliography

Fredericksen, R.E., Verstraten, F.A.J. and Van de Grind, W.A., (submitted-a). Temporal integration of random dot apparent motion information in human central vision. Submitted.

Fredericksen, R.E., Verstraten, F.A.J. and Van de Grind, W.A., (submitted-b). An analysis of the mechanism underlying temporal integration of motion information. Submitted.

Fredericksen, R.E., Verstraten, F.A.J. and Van de Grind, W.A., (in press), Spatio-temporal characteristics of human motion perception, to appear in *Vision Research*.

Lappin, J.S. and Bell, H.H. (1976) The detection of coherence in moving random-dot patterns. *Vision Research*, **16**, 161-168.

Mikami, A., Newsome, W.T., and Wurtz, R. H., (1986). Motion selectivity in macaque visual cortex. I. Spatiotemporal range of directional interactions in MT and V1. *Journal of Neurophysiology*, **55:6**, pp. 1308-1327.

Mikami, A., Newsome, W.T., and Wurtz, R. H., (1986) Motion selectivity in macaque visual cortex. II. Mechanisms of direction and speed selectivity in extrastriate area MT. *Journal of Neurophysiology*, **55:6**, pp. 1328-1339.

Scobey, R.P. and Kan, P.L.E. (1991) A horizontal stripe of displacement sensitivity in the human visual field. *Vision Research*, **31**, 99-109.

Van Doorn, A.J., and Koenderink, J.J. (1982) Spatial properties of the visual detectability of moving spatial white noise. *Experimental Brain Research*, **45**, 189-198.

Van Doorn, A.J., and Koenderink, J.J. (1982) Temporal properties of the visual detectability of moving spatial white noise. *Experimental Brain Research*, **45**, 179-188.

Van Doorn, A.J., and Koenderink, J.J. (1984) Spatiotemporal integration in the detection of coherent motion. *Vision Research*, **24**, 47-53.

Van de Grind, W.A., Koenderink, J. J. and Van Doorn, A.J., (1986) The distribution of human motion detector properties in the monocular visual field. *Vision Research*, **26**, 797-810.

Van de Grind, W.A., Koenderink, J. J., Van Doorn, A.J., Milders, M.V., and Voerman, H. (accepted) Inhomogeneity and anisotropies for motion detection in the monocular visual field of human observers. To appear in *Vision Research*.

Van de Grind, W.A., Koenderink, J.J., and Van Doorn, A.J. (1987)
Influence of contrast on foveal and peripheral detection of coherent motion in moving random-dot patterns. *JOSA A*, 4, 1643-1652.

Chapter 6: The architecture of the motion detection system

6.1 Introduction

Chapter 2 presented psychophysical data on the visibility of apparent motion and the dependence of that visibility on the position of the stimulus in the visual field. The form of the results indicates that the visibility of the spatial displacement size ($\mathbf{S_S}$) of a random dot motion stimulus depends on both the image frame duration ($\mathbf{T_S}$, or the inverse of the frame rate) and the eccentricity of the stimulus in the visual field (**E**). One especially interesting behavior of that dependence is that the maximum and minimum visible spatial displacement sizes scale differently as the stimulus position is moved farther from the fixation point (fovea). A mathematical description of this scaling difference was derived based on known neuroanatomy and neurophysiology (see appendix C of chapter 2). The description is based on the projection of visual space to the visual cortex in combination with a distribution of neuronal interconnection distances in visual cortex. In the first part of this chapter I refine the model in chapter 2 using the psychophysical data from chapters 2 and 5. The refined model accounts for (1) the differential scaling behavior of the minimum and maximum visible spatial displacement sizes with **E**, and (2) the shape of the threshold surfaces presented in chapter 2. The model describes the distribution of the distances between receptive field centers (that distance is called $\mathbf{S_d}$ in chapter 2) as resulting from the hypothesized neuronal interconnection distribution.

However, the dependence of the visible range of displacement sizes on $\mathbf{T_S}$ and the change of that range with E is not accounted for in the model. One possible explanation of this dependence comes in the form of self-organization principles. Self-organization is a common phenomenon in biological systems. The development of important pieces of biological machinery is often guided by sensory input and what appear to be some simple connectivity-creating rules. Connections between important neural areas can be formed or modified based on correlated activity between neurons. The motion detection system is formed at least in part from such a process. In the second part of this chapter I present the results of some self-organization simulations designed to explore this possibility. The self-organization model is based on psychophysical and neurophysiological data and concentrates on the formation of the neuronal interconnection distribution. The simulation explores the influence of spatio-temporal response properties of the bi-local detector's receptive fields on the self-organization process (see figure 2.1).

Before describing the details of the models presented here, I will briefly review important results and symbols. Luminance signal-to-noise ratio (LSNR) thresholds for directional motion were measured over a range of $\langle \mathbf{S_S}, \mathbf{T_S} \rangle$ combinations and at several positions in the visual field (see figures 2.3 and 2.4) . The resulting threshold values represent a threshold surface over the $\langle \mathbf{S_S}, \mathbf{T_S} \rangle$ plane (figure 2.2). Previous measurements of the limits of perceivable spatial displacement sizes used a different measurement paradigm: $\mathbf{S_S}$ was varied to find an absolute threshold level (usually at either 71% or 79% correct). In order to compare the LSNR threshold measurements to the previous data, the LSNR = 100 surface contour was chosen as a fiducial point. The portion of this contour on the lower $\mathbf{S_S}$ side of the surface is referred to as S^{100}_{min} and the portion of the contour on the upper $\mathbf{S_S}$ side of the surface is referred to as S^{100}_{max}. These contours are marked in figures 2.3 and 2.4 by the thicker contour lines.

6.2 Model description

As a summary of the differential dependence of S^{100}_{min} and S^{100}_{max} contours on **E** and $\mathbf{T_S}$, those contours are extracted from the data of subject FV (from figures 2.3 and 2.4) and shown together in figure 6.1. The contours are taken from threshold surfaces measured at **E** = 0 and **E** = 24 degrees.

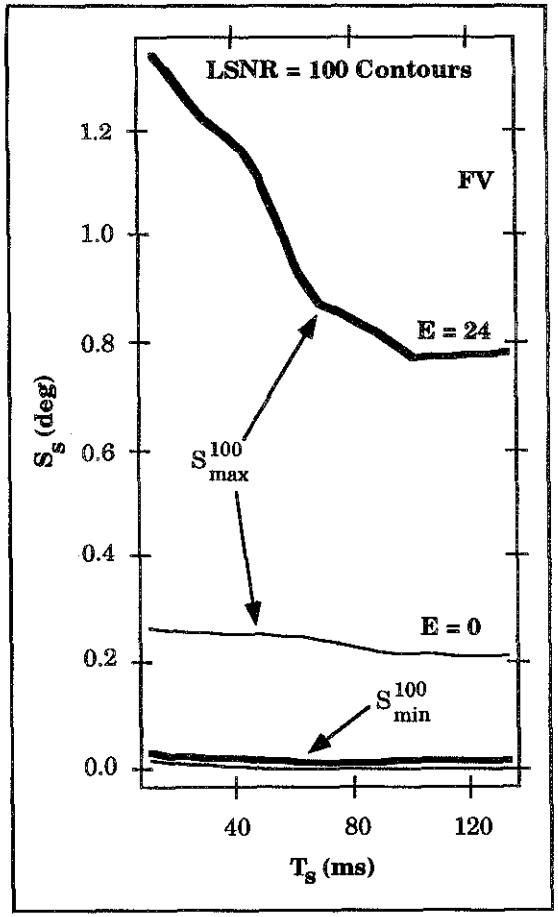


Figure 6.1
The LSNR = 100 contours are plotted for subject FV from figures 2.3 and 2.4. The thin lines are taken from the **E** = 0 results and the thick lines are taken from the **E** = 24 results.

Figure 6.1 shows clearly the difference in scaling of S^{100}_{min} and S^{100}_{max} with **E** as well as the dependence of that scaling on $\mathbf{T_s}$. The structure of the model that is presented in the first half of this chapter explains the dependence of scaling on **E** while discussion of the origin of the $\mathbf{T_s}$ dependency is deferred to the second part of the chapter. A derivation of a mathematical description of the E-scaling property of the S^{100} contours is given in appendix C of chapter 2. Figure 6.2 shows a schematic representation of visual anatomy used to derive the mathematical relationship. The figure is a high level description[1] of visual anatomy and includes a representation of the retina, the lateral geniculate nucleus, and visual cortex.

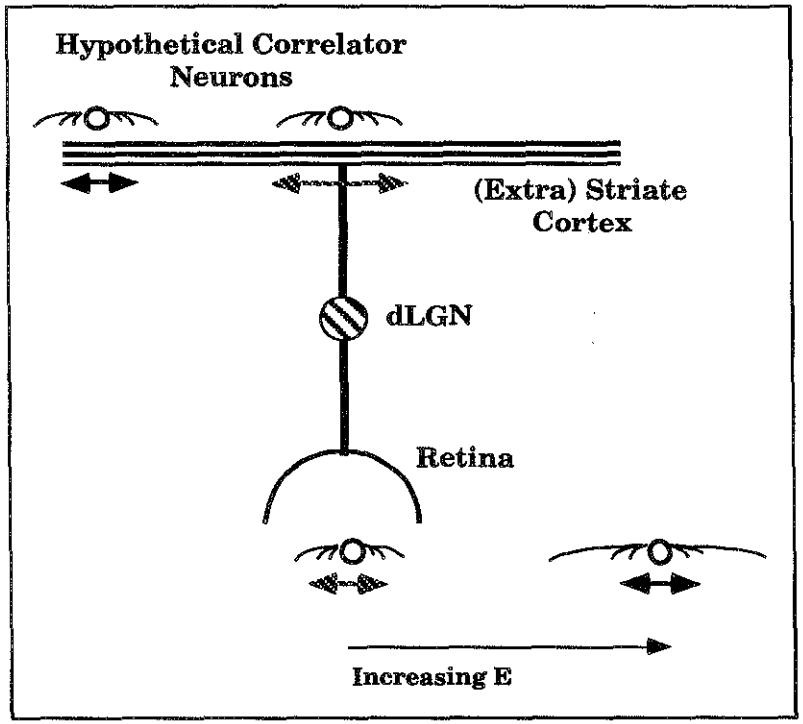


Figure 6.2
A high level schematic of a primate visual system is shown. The arrows in front of the retina represent sizes of objects in the visual field. The arrows below the cortex represent the size of the cortical representation for each arrow in the visual field. The "spiders" in the figure over the cortex represent the neurons that implement the correlation, or motion calculation function. The spiders over the arrows in visual space demonstrate the distance in visual space that is covered by each neuron in cortex.

[1] Discussion of visual system anatomy and physiology in this chapter is directed specifically toward the primate visual system unless stated otherwise.

The retina projects, via the lateral geniculate nucleus, to striate cortex and then beyond (extra-striate areas). The mapping property of interest here is the characteristic projection of the retinal ganglia (and hence visual space) to area V1 of visual cortex (and beyond).

The initial mapping of visual space to area V1 (the first high level stage of visual processing) is often described in the literature by

$$M = \frac{M_f}{\left(1+\frac{E}{E_2}\right)} \qquad \text{mm/deg} \tag{6.1}$$

where **M** is called the cortical magnification factor. $\mathbf{M_f}$ is the magnification of visual space (in millimeters per degree) at the fovea, and $\mathbf{E_2}$ is the angular distance from the fovea at which the magnification falls to one half that in the fovea. **M** provides an estimate of the size of a cortical representation of 1 degree of visual space at a distance **E** away from the fovea. This equation has been fit to a number of perceptual phenomena (e.g. see Drasdo, 1991 for a review) including retinal ganglion cell receptive field size and visual acuity. Fitting the equation to other perceptual phenomena requires selection of $\mathbf{M_f}$ and $\mathbf{E_2}$.

We must consider where the percept of motion arises when applying this equation to the perception of motion. We know that motion sensitivity properties arise in area V1 and are transmitted to and refined by later stages of cortical processing. These higher level visual areas generally retain the spatial mapping characteristics (except perhaps at higher stages of visual processing, e.g., area MT). The point at which our percept of motion actually begins is unclear. The projection of visual space that we are considering must therefore be considered as a composite, or aggregate of several processing layers.

The influence of the mapping is indicated by the arrows and "spiders" in figure 6.2. Constant-sized arrows in the visual field are shown mapped to their cortical representations. The correspondence of the arrows is indicted by their shading. The arrow directly in front of the retina represents the visual extent of an object at the fovea. The arrow is magnified with respect to other objects when it is mapped to cortex. Arrows that are farther away from the fovea are mapped to smaller distances in cortex than arrows near the fovea. The arrow farthest to the right (farthest from the fovea in the figure) in the figure has the smallest representation in cortex.

The correlation operation (speaking in terms of the bi-local detector model as described in chapter 2) must take place in visual cortex because directional selectivity is not present in lateral geniculate nucleus cells. The "spiders" placed above the representation of cortex in figure 6.2 represent a hypothetical set of neurons that perform this correlation function. They are shown as being of uniform size and extent (consistent with known cortical architecture). The legs of the "spiders" represent axonal or dendritic arborizations of the neurons. That arborization represents the cortical distances over which a neuron can gather information for correlation. I will call the distribution of those distances $P(\mathbf{D_c})$. The correlation neurons are projected back to visual space in figure 6.2 as an alternative but equivalent display of the influence of cortical magnification. Neurons farther away from the foveal representation in cortex are able to correlate information over larger distances in visual space.

The architecture proposed in chapter 2 included a $P(\mathbf{D_c})$ that is constant[2] across visual cortex (and hence with **E**). The possible correlation distances in cortex are mediated by $P(\mathbf{D_c})$, but the visual spatial distances that $P(\mathbf{D_c})$ covers is a function of the mapping of visual space to cortex described by **M**. Equations describing the approximate mapping of visual spatial distances to cortical distances can be derived based on **M**. If we select two points on a radial line through the fovea, we can use equation (6.1) to calculate the distance in cortex between the cortical representations of the two points. Equation (6.1) can be viewed as the instantaneous derivative of cortical distance with respect to visual distance. Integration of equation (6.1) with respect to **E** gives us a mathematical relationship between a distance in visual space and a distance in visual cortex. The detailed derivation of that relationship is

[2] For now we assume that $P(\mathbf{D_c})$ is constant across visual cortex because other cortical structures related to vision change little across cortex, e.g. ocular dominance columns.

presented in appendix C of chapter 2. For convenience we repeat the final results here. A distance in visual space that begins at a position **E** on the previously mentioned radial line through the fovea and ends a distance $\mathbf{S_s}$ further along that radial line is mapped to a distance $\mathbf{D_c}$ in visual cortex described by

$$(6.2) \qquad D_c = M_f \cdot E_2 \cdot \ln\left(\frac{E_2 + E + S_s}{E_2 + E}\right) \quad mm$$

The inverse mapping (solving equation (6.2) for $\mathbf{S_s}$) is

$$(6.3) \qquad S_s = (E_2 + E) \cdot \left(e^{\left(\frac{D_c}{M_f \cdot E_2}\right)} - 1\right) \quad deg$$

The variables $\mathbf{M_f}$, $\mathbf{D_c}$, **E**, $\mathbf{E_2}$, and $\mathbf{S_s}$ are as defined above. Together the two equations describe the relationship between points in visual space and points in visual cortex as described by the arrows and spiders in figure 6.2. The distribution $P(\mathbf{D_c})$ is assumed to be constant across the cortex. However, the range of visual spatial distances represented by $P(\mathbf{D_c})$ changes with **E** because of **M**. The equations therefore provide some insight into how the visible range of visible spatial displacements changes with **E**.

6.3 Modeling psychophysical data

The next question that must be answered is how well psychophysical data follows equations (6.2) and (6.3). I first compare the qualitative performance of the mapping to some results previously reported. Baker and Braddick (1985) investigated the change of the limits of perceivable spatial displacement across **E** using a somewhat different procedure than that used here. They used two-frame apparent motion to measure the minimum and maximum visible displacements in the visual field. The two measures are commonly referred to as D_{min} and D_{max}. Chapters 2-5 argue that two frame apparent motion does not provide an optimal stimulus duration for measuring these quantities. However, their data show an important qualitative property described by equation (6.3). In terms of $P(\mathbf{D_c})$ the values of D_{min} and D_{max} can be considered to map to the largest and smallest interconnection distances available in cortex. If we fix $\mathbf{D_c}$ ($\mathbf{M_f}$ and $\mathbf{E_2}$ are already constants), then equation (6.3) reduces to a linear equation in **E**.

$$(6.4) \qquad S_s = K_c \cdot E_2 + K_c \cdot E \quad deg$$

where

$$(6.5) \qquad K_c = \left(e^{\left(\frac{D_c}{M_f \cdot E_2}\right)} - 1\right)$$

Equation (6.4) describes a linear dependence of $\mathbf{S_s}$ on **E** for a fixed value of cortical distance $\mathbf{D_c}$. It is clear that the value of $\mathbf{D_c}$ can have a great influence on $\mathbf{K_c}$, the slope of this linear relationship, because it occurs as an exponent in equation (6.5). Figure 6.3 shows a representation of data from Baker and Braddick (1985) for values of D_{min} and D_{max} in the first ten degrees of visual angle from the fovea. When plotted together, the slopes of the lines fitted to the values of D_{min} and D_{max} are drastically different. The slopes of D_{min} and D_{max} with respect to **E** for subject CLB (see their figures 3 and 4) are, fitting the published data by hand, 0.002 degrees/degree ($\mathbf{S_s}$/E) and 0.19 degrees/degree, respectively. The ratio of the slopes is about 95 and is consistent with equation (6.5).

It is also interesting to apply this model to the psychophysical data presented in chapters 2-5. To do so, we must fix some parameters in the mathematical description. Estimates of the value of $\mathbf{M_f}$ in primates (e.g. see Drasdo, 1991 for a review) based on spacing of retinal ganglia and/or projections of retinal ganglia to area V1 has ranged from 8 to 22 mm/deg. A conservative estimate is 10 mm/deg, and we will use that value for the current discussion.

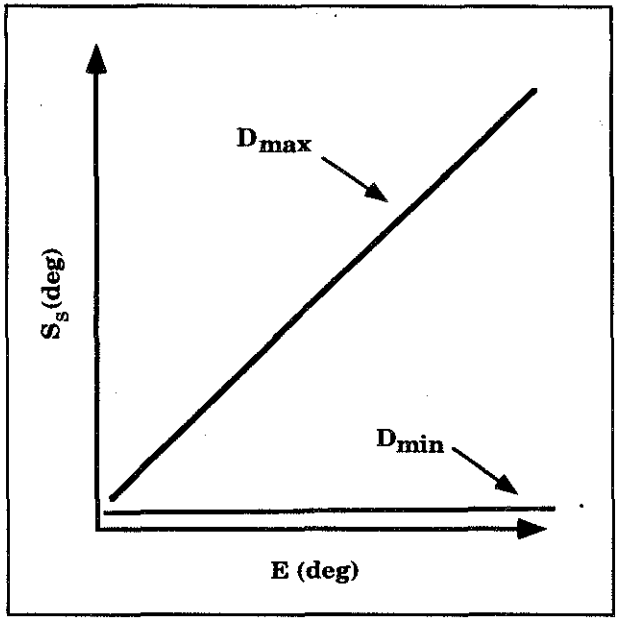


Figure 6.3
A representation of data from Baker and Braddick (1985) for values of D_{min} and D_{max} is shown. The data were measured in the first ten degrees of visual angle from the fovea. When plotted together the slopes of the lines fitted to the values of D_{min} and D_{max} are drastically different.

We next need to account for $\mathbf{E_2}$ in the relationship. Calculation of this value from our data (figure 2.5) show a dependence of $\mathbf{E_2}$ on $\mathbf{T_s}$. However we must be careful here because that value will depend on the stimulus dimensions and the scaling of those dimensions with **E**. The value of $\mathbf{S_s}$ and the dimension of the stimulus parallel to the direction of $\mathbf{S_s}$ determine the ratio of correlated to uncorrelated pixels, or the spatial signal-to-noise ratio (SSNR). The influence of spatial noise has been discussed in appendix A of chapter 2 and in chapter 5. Although the consistent performance of $\mathbf{E_2}$ across subjects as shown by figure 2.5 is important, the actual calculated values of $\mathbf{E_2}$ are only good for our own stimulus measurement method: changing the stimulus dimensions or the scaling of the stimulus dimensions with **E** will change the SSNR and therefore the quantitative values of $\mathbf{E_2}$ in figure 2.5. With that caveat in mind we fit the average of the curves in figure 2.5 using splines and then use that approximation as $\mathbf{E_2(T_s)}$ in our model[3]. The model is therefore specific to our measurement paradigm (and therefore can not be applied to the data on D_{min} and D_{max}).

The next step is to select an appropriate form for $P(\mathbf{D_c})$. The distribution should be unimodal and always positive. A convenient function for this is a Gamma distribution because it has two parameters that can be used to tune the shape of the curve[4]. The cortical interconnection distance distribution therefore is modeled by

$$P(D_c;\alpha,\beta) = \frac{x^{(\alpha-1)} \cdot e^{\left(-\frac{D_c}{\beta}\right)}}{\Gamma(\alpha)\cdot\beta^{\alpha}} \tag{6.6}$$

where α and β are the distribution tuning parameters and $\Gamma(\alpha)$ is the Gamma *function* (not distribution) defined as

$$\Gamma(\alpha) = \int_0^{\infty} e^{-t} \cdot t^{\alpha-1} dt \tag{6.7}$$

To fit this distribution to a subject's performance we need to select the values of α and β. Restrictions for the distribution can be obtained by inspecting the psychophysical data at a single $\mathbf{T_s}$ value. Figure 2.3 shows that the S^{100}_{min} and S^{100}_{max} values for subject FV at $\mathbf{D_s}$ = 1000 ms, $\mathbf{T_s}$ = 11 ms in the fovea are approximately 0.008 degrees and 0.25 degrees, respectively. Chapter 5 discussed the relationship between theoretical detector count and LSNR threshold: the LSNR threshold is proportional to $\mathbf{N^{-0.5}}$ where **N** is the number of detectors responding to the stimulus.

[3]We note that $P(\mathbf{D_c})$ and $\mathbf{E_2(T_s)}$ together determine the scaling characteristics of $\mathbf{S_s}$ as a function of **E**. Physiologically, however, it is possible that $\mathbf{E_2}$ is a constant and $P(\mathbf{D_c})$ is a function of $\mathbf{T_s}$. This possibility is discussed in the second part of this chapter in terms of self-organization of the system. In the meantime it is convenient to consider the model in terms of a constant $P(\mathbf{D_c})$ and varying $\mathbf{E_2}$ using the fitted spline approach.

[4]Other distributions may be equally good but I am aware of no theoretical reason to choose one function over another.

$P(D_c)$ must be very small at S_{min}^{100} and S_{max}^{100} because the subject's sensitivity is very low, or alternatively, the LSNR threshold is very high. There must therefore be very few detectors responding to those values S_s values. Another restriction on α and β is that the peak value of the distribution should correspond to the minimum threshold values seen in the psychophysical data. For subject FV (for the same curve as referenced above) that optimum is between $S_s = 0.05$ and $S_s = 0.1$. We will refer to the values of D_c that correspond to S_{min}^{100} and S_{max}^{100} as D_{min}^{100} and D_{max}^{100}, respectively (calculated using equation (6.2), E, and $E_2(T_s)$). The values of α and β were obtained by a least squares solution for the above mentioned threshold curve for subject FV with the constraints

$$
\begin{aligned}
P\left(D_{min}^{100};\alpha,\beta\right) &= 0.006 \\
P\left(D_{max}^{100};\alpha,\beta\right) &= 0.006 \\
\mathcal{E}\left[P\left(D_c;\alpha,\beta\right)\right] &= 0.87
\end{aligned}
\tag{6.8}
$$

where $\mathcal{E}[]$ represents the expected value operator and provides the mean of the distribution[5]. The resulting values of α and β were (to three decimal places) 5.083 and 0.171, respectively. The distribution for these parameter values is shown in figure 6.4.

A maximum interconnection distance of about 2.5 mm in cortex is not at all unreasonable given that other cortical processing architectures such as ocular hypercolumns in primates and man are of the same order of magnitude in size. Moreover, there are intrinsic connections that run horizontally through cortex as far as 10 mm (Luhmann, Martínez Millán, and Singer, 1986).

[5] A mean D_c value of 0.87 maps to a higher S_s value than the optimum S_s value in subject FV's data but the peak of the distribution is at a lower D_c value than the mean. A mean of 0.87 brings that peak to an approximately correct position for the psychophysical data.

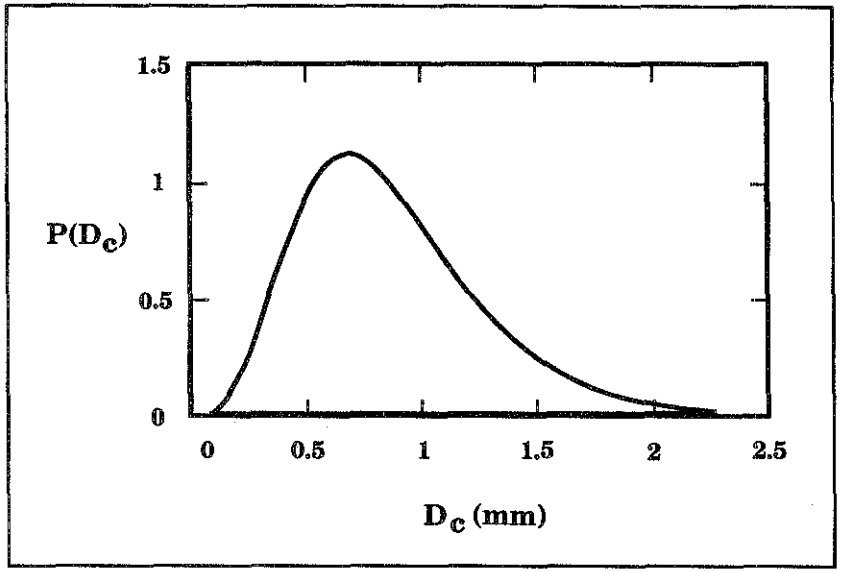


Figure 6.4
The threshold of subject FV for T_s = 11 ms, D_s = 1000 ms, E = 0 (figure 2.3) were used to fit a hypothetical but corresponding cortical interconnection distance distribution, $P(D_c)$, using the gamma distribution. The fitted curve is shown. See text for an explanation of the distribution form.

The goal of fitting this distribution is to apply the refined model to the psychophysical data and see how the results correspond with the rest of the psychophysical data. To continue we must consider how to compute from the model a threshold value that is equivalent to the psychophysical threshold values. Chapter 5 discusses the relationship between the number of detectors, the stimulus height and width, and the resulting perceptual threshold. That same inverse square root relationship applies here. However, we must also include the finite area of the stimulus display. Equation (6.6) gives us a point-by-point density of detectors: to relate this to psychophysical results, we need to integrate over the area of the stimulus. We must take care to include the spatial displacement size in relation to the width of the screen because the effective width is reduced by S_s. The equation for the total number of detectors for a displacement size S_s at a frame duration of T_s with a stimulus centered at E_c and width W is

$$(6.9)\quad N(S_s, T_s, E_c, W) = \int_{E_c - \frac{W}{2}}^{E_c + \frac{W}{2} - S_s} P\left(M_f \cdot E_2(T_s) \cdot \ln\left(\frac{E_2(T_s) + E + S_s}{E_2(T_s) + E}\right)\right) dE$$

The presence of α and β as parameters to $P(\mathbf{D_c})$ are assumed for the rest of the chapter.

6.3.1 Comparison with $<S_S,T_S>$ threshold surfaces

If we evaluate N() in equation (6.9) for the given parameters, we can use equation (5.6) to plot a threshold (i.e., plot the value $N^{-0.5}$) for a range of parameters. Figure 6.5 shows the outcome of such an evaluation at $\mathbf{E} = 0$ and $\mathbf{E} = 24$ degrees as contour diagrams of surfaces across $\mathbf{T_S}$ and $\mathbf{S_S}$. The hypothetical screen size was scaled according to the method used for the psychophysical measurements in chapter 2. The two contour diagrams correspond to the contour diagrams of the psychophysical thresholds in figures 2.3 and 2.4 for subject FV.

Comparison of the model figures with subject FV's psychophysical thresholds shows a good bit of similarity. This is not too surprising since the model was fit to the subject's data. However, doing the comparison allows us to confirm that the whole model works even though we fit only the data for the $\mathbf{T_S}$ = 11 ms threshold curve at the fovea.

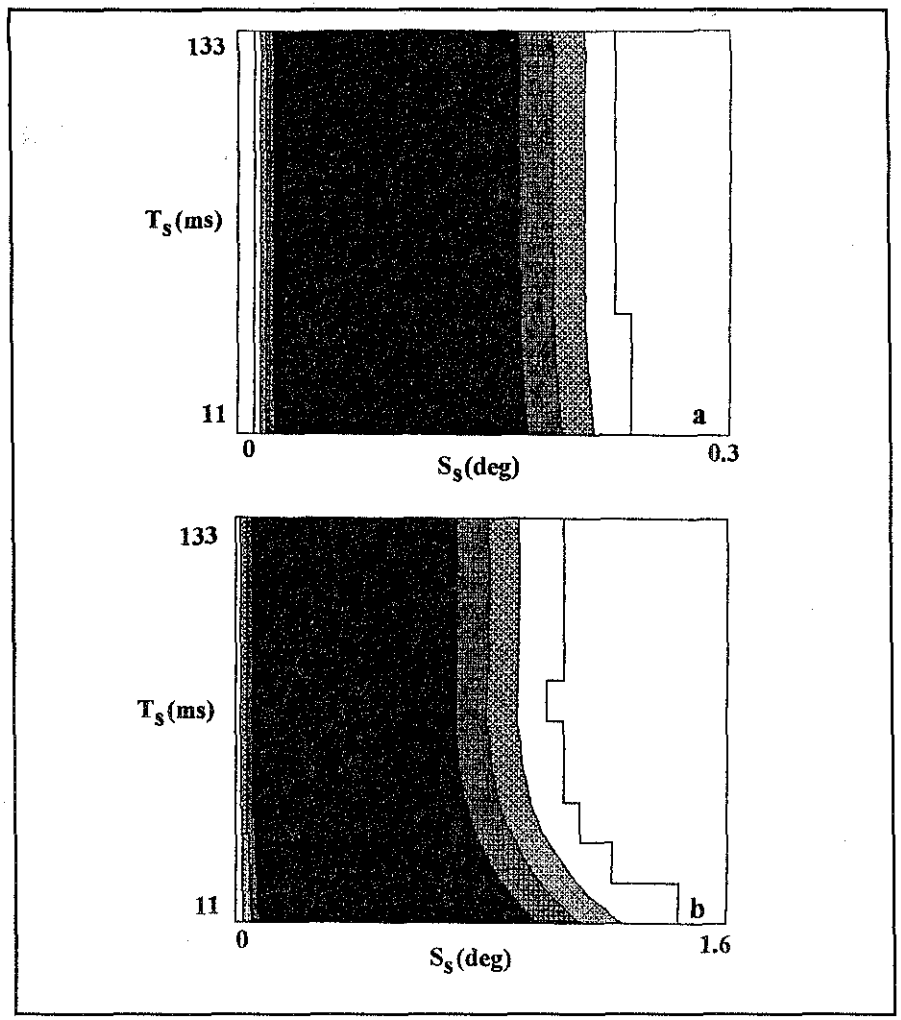


Figure 6.5
The outcome of an evaluation of equation (6.9) is shown for parameter values of $\mathbf{E} = 0$ (figure a) and $\mathbf{E} = 24$ degrees (figure b) as contour diagrams of surfaces across $\mathbf{T_S}$ and $\mathbf{S_S}$. The model threshold values were calculated according to probabilistic summation. The hypothetical screen size was scaled according to the method used for the psychophysical measurements in chapter 2. Figures a and b correspond to the psychophysical thresholds in figures 2.3 and 2.4, respectively, for subject FV.

6.3.2 Comparison with spatial summation curves

We can now compare the fitted model to the spatial summation experiments in chapter 5. However, I first discuss the relationship of $P(\mathbf{D_C})$ to the distribution of $\mathbf{S_S}$ values in the visual field. In equation (6.9), a single value of $\mathbf{S_S}$ is mediated by different cortical distances depending on **E**. That means that the density of connections (or detectors) mediating that single spatial displacement size in the visual field changes across **E** according to $P(\mathbf{D_C})$. I will refer to that density as $Q(\mathbf{S_S},\mathbf{E})$. Figure (6.6a) shows the form of $Q(\mathbf{S_S},\mathbf{E})$ for several values of $\mathbf{S_S}$ across **E**. The curves were generated using the fitted distribution in figure 6.4 and the same $\mathbf{S_S}$ values as in figure 5.4. $P(\mathbf{D_C})$ is symmetric about the fovea, and the model 8-degree-wide stimulus is centered on the fovea, so the curves are plotted from 0 to 4 degrees in the graph. The coarsely dotted line represents $\mathbf{S_S}$ = 0.031 degrees, the finely dotted line represents $\mathbf{S_S}$ = 0.094 degrees, and the solid line represents $\mathbf{S_S}$ = 0.0188 degrees. It is clear that the density of detectors for each displacement size is different. It is also clear that the density varies across the area covered by the 8 degree wide stimulus. This variation can influence the resulting spatial summation properties for various sized displays.

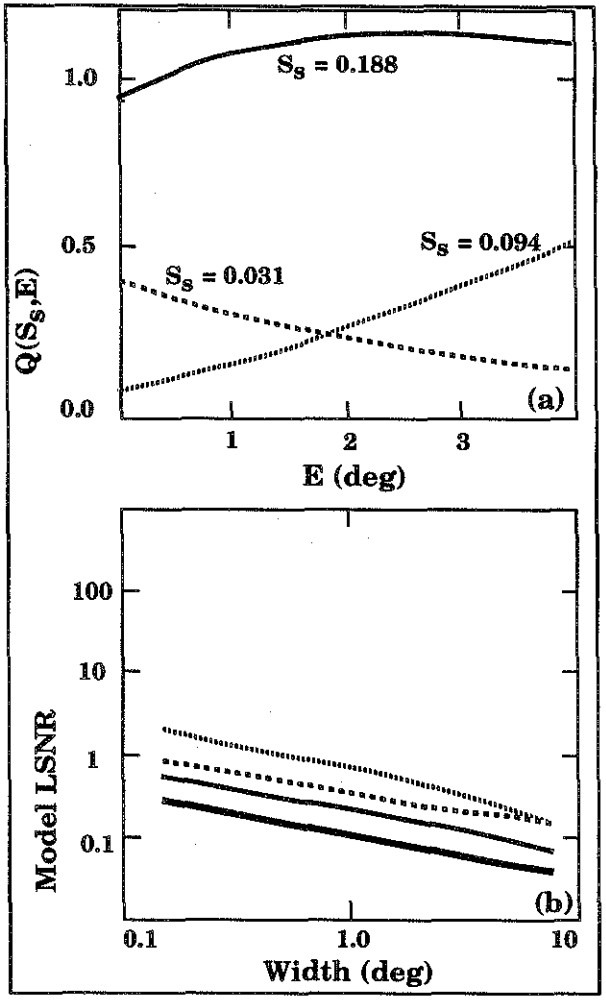


Figure 6.6
(a) The density distribution of visible $\mathbf{S_S}$ values mediated by $P(\mathbf{D_C})$ are plotted across the hypothetical display as a function of **E**. The coarsely dotted line represents $\mathbf{S_S}$ = 0.031 degrees, the finely dotted line represents $\mathbf{S_S}$ = 0.094 degrees, and the solid line represents $\mathbf{S_S}$ = 0.0188 degrees. (b) Equation (6.9) was evaluated following the spatial summation experiments in chapter 5. Parameter values of **E**=0 and $\mathbf{T_S}$ = 33 ms were used while **W** was varied from 0.1 to 8 degrees. The $\mathbf{S_S}$ values are as in (a).

For a fixed stimulus width **W**, we can evaluate the number of detectors under the one-dimensional model stimulus using equation (6.9). We then use the $N^{-0.5}$ relationship described in chapter 5 to calculate a model threshold value. To compare most directly to the data in figure 5.4, we set **E**=0 and $\mathbf{T_S}$ = 33 ms and then calculate a model threshold for values of **W** from 0.1 to 8 degrees. This produces the curves shown in figure 6.6b. The extra thick line at the bottom is a reference line representing the inverse square root law. The same $\mathbf{S_S}$ values were used as in figure 6.6a, and each $\mathbf{S_S}$ value was plotted in the same line style as in figure 6.6a for easy comparison. $\mathbf{Q(S_S,E)}$ is not linear but the model spatial summation threshold curves show a surprising adherence to the reference curve.

Adherence to the probabilistic summation law only requires an approximately linear behavior of $\mathbf{Q(S_S,E)}$ with **E**. Some influence of the non-linear nature of the distribution is shown at the largest display widths, but the degree of deviation is small compared to the measurement noise in the actual psychophysical data. We cannot therefore point to corresponding non-linear behavior in the psychophysical data in chapter 5. However, the model would predict that this non-linearity would become more prominent for a wider stimulus.

6.3.3 Model predictions

The model can be verified by, or provides predictive capabilities for, future threshold measurements. For example, if the model is accurate, we should be able to predict the range of perceivable $\mathbf{S_S}$ values for various display-width scaling strategies (visual field extent of the stimulus). Another possible test of the model would be the change of perceivable $\mathbf{S_S}$ values if the display width is kept constant and the display is placed at various positions in the visual field. In addition, the non-linear behavior of $\mathbf{P(D_C)}$ as evidenced in the spatial summation curves in figure 6.6 can be tested by using the appropriate display dimensions in an actual psychophysical experiment.

6.4 Self-organization simulations

There is an interesting characteristic of the psychophysical threshold surfaces that is not covered by the model. The psychophysical threshold surfaces in chapter 2 show variations in S^{100}_{min} and S^{100}_{max} with $\mathbf{T_S}$ that correspond to what must be real variations in $\mathbf{P(D_C)}$: S^{100}_{min} and S^{100}_{max} are not truly constant at the fovea, nor do they approach a constant value even for large $\mathbf{T_S}$ values outside the fovea. These variations are good evidence that a constant $\mathbf{P(D_C)}$ provides only a first order approximation to the distribution of perceivable $\mathbf{S_S}$ values. Chapter 2 discusses the possible role of self-organization as a basis for the distribution of visible $\mathbf{<S_S,T_S>}$ combinations across the visual field. In this section I present the results of simulations designed to explore that possibility.

6.4.1 Motivation

The simulation design is based on the architectural scheme presented in figure 6.2. It is assumed that the correlator neurons in visual cortex must select some set of receptive fields as inputs through a self-organization process. (NOTE: for the remainder of this section an unqualified reference to receptive fields refers to either LGN or bi-local detector input receptive fields as indicated by the context of the discussion. References to the composite receptive field of the motion detector will be specifically indicated as such.)

Common rules of self-organization processes are based on selecting inputs that are correlated. This means that the spatio-temporal properties of the receptive field components that are available to build a motion detector in combination with the visual input to the system during the self-organization process should influence the range of perceivable $\mathbf{<S_S,T_S>}$ combinations emerging from that self-organization process. If a particular range of spatio-temporal information (we can think of that as a range of $\mathbf{<S_S,T_S>}$ values) is not present during self organization, we would not expect the emergent array of motion detectors to be able to respond to that range of motion information. Likewise we would not expect the motion detector array to detect $\mathbf{<S_S,T_S>}$ combinations that were available during the array formation process but could not activate the available pool of receptive fields. In short, the matchup of motion information and the available processing equipment should influence the properties of the motion detectors that result from self-organization. The simulation was designed to explore this hypothesis.

6.4.2 Design

The simulation results are most valuable if they have some relation to actual neurophysiology and anatomy. To achieve this we need to take a few guesses at the physical locations of some of the components of the motion detectors and then use the known properties of the actual physical components in the simulation. We must also select rules for the self-organization process and the structure of the simulation. Each of these choices is discussed below.

6.4.2.1 Receptive field response

The directionally selective neurons in V1 must be choosing the outputs of lateral geniculate nucleus cells (or some partially processed version of their outputs) as inputs to correlate in some fashion. The form of that correlation is not a question here. Nor is the question whether V1 is a basis for a physical percept. It is certain, however, that in primates the first directionally selective neurons in the primary visual pathway begin in area V1. I have therefore designed the receptive field responses using available literature on the spatio-temporal properties of lateral geniculate nucleus cells.

The spatial center-surround nature of lateral geniculate nucleus (LGN) receptive fields is well known. Because the off-center/on-surround cells are essentially contrast-preference reversed versions of the on-center/off-surround cells, only on-center cells style are used in this simulation. I use the common approach of modeling that spatial receptive field as a balanced difference-of-Gaussian (DOG) spatial filter kernel. Less is known about quantitative measures of the temporal response of the LGN cells, but some interesting results have appeared in the last few years. An experimental paradigm first used by Mastronarde (1987a,b) has revealed that there are apparently two classes of X-type LGN cells in cat corresponding to lagged and non-lagged temporal response properties. The lagged cells form a population with responses to visual input that are delayed in time in relation to the non-lagged cells to such an extent that the two classes can be easily classified by that lag time. Saul and Humphrey (1990) have extended this result to the Y-type LGN cells as well, showing that there appear to be both lagged and non-lagged X and Y cells and that each of those four populations has a range of fast-transient (FT) and slow-sustained (SS) type response profiles (outside of the absolute delay). It is extremely tempting to identify these temporal response properties as a physiological basis for the bi-local style motion detector: the lag may correspond to the differential delay of the receptive field outputs in the bi-local detector, and the range of FT and SS type response profiles could serve as a basis for detecting various visual velocities.

We must be careful in suggesting these correspondences as truth because the neurophysiological data is new and must still stand the test of time. Nevertheless, the results provide an intriguing basis for the simulation design. More importantly, Saul and Humphrey (1990) present linear filter impulse response functions for the four main classes of lagged and non-lagged FT and SS LGN responses. The linear filters are approximated using the DOG technique (in time), so creating the spatio-temporal filter response for the simulation is straightforward. There exists, of course, the justifiable objection that the response properties of the LGN cells are not linear. However, the linear portion of the filter response captures the qualitative temporal response properties of LGN cells. Moreover, the self-organization part of the simulation is based on the solutions to a set of coupled differential equations and is highly abstracted from the actual biological organization process. The realism of the simulation is therefore limited by the self-organization rules rather than by the failure to model non-linearity of the LGN cell temporal response.

The simulation was designed for one dimension of space plus time in order to make it more manageable. The spatio-temporal response of the LGN was a spatio-temporally separable filter impulse response constructed using the DOG method in both dimensions. The surround (negative Gaussian) part of the spatial filter always had a standard deviation of 1.7 times that of the center (positive) Gaussian. The temporal impulse response of the filters were created by approximating the DOG-in-time filters shown in the paper by Saul and Humphrey. All motion inputs and filters were represented as 2-dimensional space-time arrays

(images). Figure 6.7 shows an example of the filter kernel as a two dimensional surface over the space-time plane.

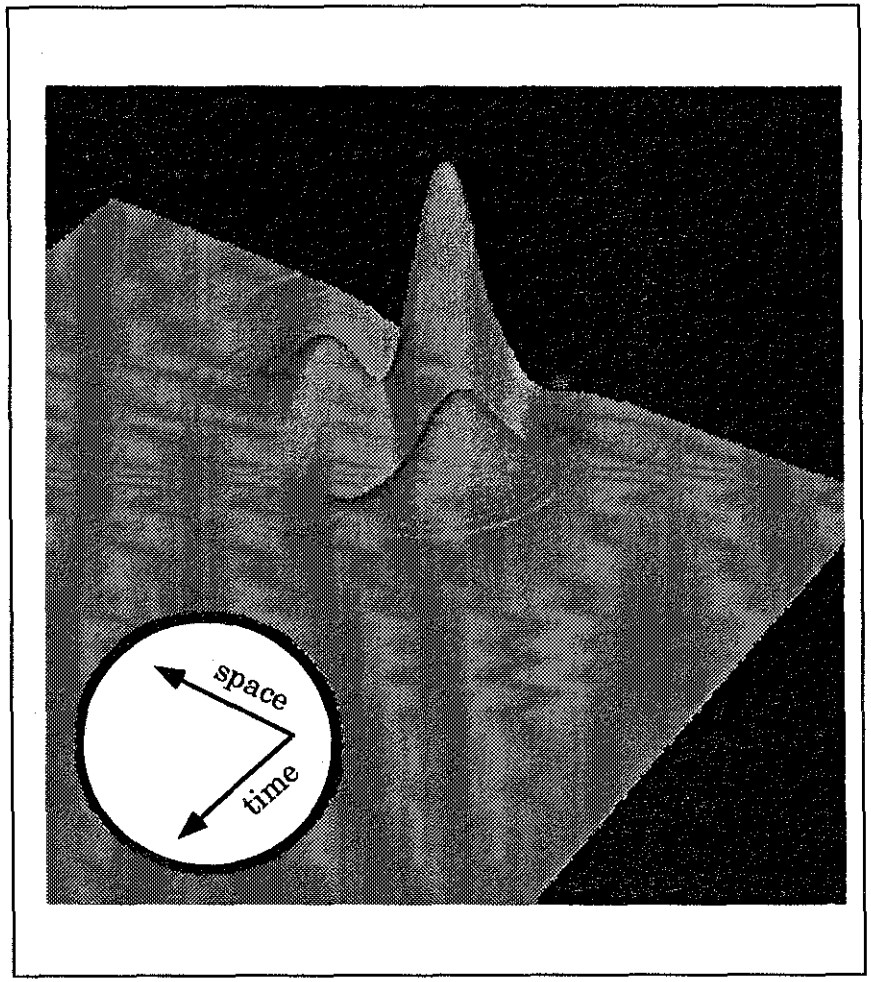


Figure 6.7
An example LGN-response-like spatio-temporal filter is shown as a two dimensional surface over the space-time plane.

The use of spatio-temporal separable filters is not realistic also because there is evidence of spatio-temporal inseparability in the receptive fields of retinal ganglion cells (Fleet and Jepson, 1985) and in directionally selective neurons in cortex (Reid, Soodak and Shapley, 1991; Emerson, Bergen and Adelson, 1992). The differences in performance due to the separability or inseparability of the receptive field are important but not the issue in the current simulation. In cortical motion detectors that issue is tied up in the form of the spatio-temporal correlation. The simulation presented here is only designed to explore which receptive fields may be selected for use in that correlation process. The retinal ganglion cell or LGN cell response inseparability should perhaps be considered but is easily subsumed in the question of the available range of response types. The space-time separability issue is therefore neglected in the simulation.

6.4.2.2 Motion input

The next step in the simulation process requires selection of a motion input for driving the self-organization process. Two possibilities that were considered were (1) random motion noise represented by spatially and temporally uncorrelated random dot images and (2) motion information corresponding to egomotion, or movement of the viewpoint through the world. There are good arguments for both methods, but preliminary tests of the artificial LGN receptive fields on both types of images suggest that the choice is not important. The spatio-temporal filters that represent the artificial LGN responses produce approximately equivalent results when applied to either type of input. Although I did not test that assumption extensively, I decided that use of the egomotion style of input was more realistic. Attributing a sense of realism to the simulation with respect to the motion input depends on the time course of the development process of motion perception. If most of the development process occurs before the child is born, we would expect that random motion noise might be more appropriate. However, there is evidence that at least some amount of development of the motion detection system occurs after birth (Wattam-Bell, 1991, 1992). The duration and functional (anatomical) location of that development must be considered, but a baby is in any case moved through the world by various means thereafter either by its parents or by self-locomotion. For this reason I chose to employ motion inputs modeled after realistic egomotion.

To produce that motion information, I used the observer-world situation shown in figure 6.8. The figure describes egomotion in terms of a monocular (in accordance with our measurement of monocular psychophysical data) observer moving through the world at velocity $\bar{V}$. The important element in this situation is the velocity of the retinal images of objects in the observer's world. For an object a distance **d** from the observer we can write

$$\dot{\Theta} = \frac{\|\bar{V}\|}{d} \cdot \cos(\Theta) \text{ (deg/sec)} \tag{6.10}$$

where $\dot{\Theta}$ is the retinal angular velocity in degrees per second, Θ is the visual angle between the point of gaze and the object, and $\|\bar{V}\|$ is the magnitude of the observer's velocity.

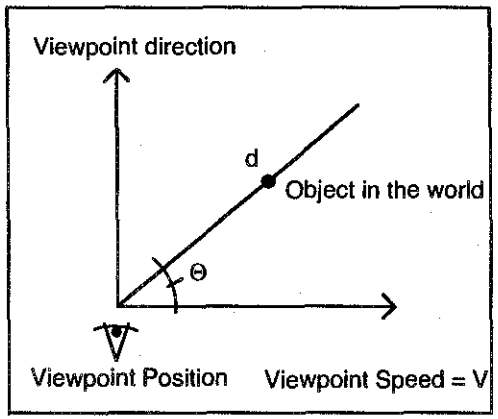


Figure 6.8
The figure describes ego motion in terms of a monocular observer moving through the world. The variable **d** is the distance from the observer to the object in the world, Θ is the visual angle between the point of gaze and the object, and $\bar{V}$ is the observer's velocity.

Equation (6.10) shows that the distribution of motion information due to ego movement changes across the visual field. That fact is important in the discussion, but the simulation is designed to be independent of Θ, so we can neglect that part of the equation for now. What is important to creating an egomotion input stimulus is that the retinal speed of an object is inversely proportional to that object's distance from the observer. Simple geometric considerations also show that the scale of the object on the retina is inversely proportional to the distance **d**. (I will use the terms scale and size interchangeably. When applied to the DOG filters, the use of the term scale implicitly refers to the standard deviation of the Gaussians.)

In our hypothetical observers world there will be some set of objects visible along any angle Θ in the visual field. There will be some distribution of sizes for these objects, but we can consider the case where all objects are the same size without any loss of generality. The distance **d** to an object governs both the extent and the angular velocity of that image on the retina. The size and angular speed of an object on the retina increases with decreasing distance. The motion stimulus used as input to the simulation was therefore a composite of random dot motion speeds and object sizes (scales) such that size and speed are proportional. Space-time images representing a range of fast and large to small and slow random dot patterns were generated and added together (luminance, or pixel-by-pixel addition) to create the composite image.

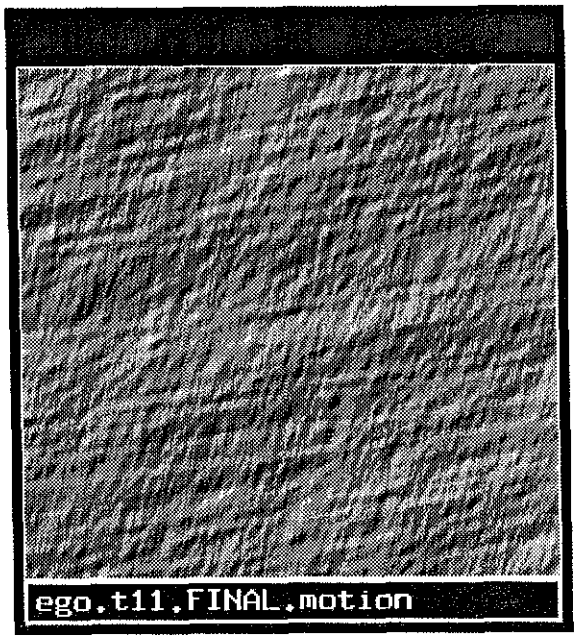


Figure 6.9
An example space-time motion image is shown. The image represents 1-dimensional random dot patterns moving from right to left. Space is the horizontal axis. Time is the vertical axis and increases from top to bottom. The speed of a pattern is indicated by the slope of the space-time path of that pattern in the image. Pixel speed and size in the image increase together.

An example composite space-time motion image is shown in figure 6.9. The movement of the random dots from right to left in the display results in luminance bars that slant from the top right to bottom left of the image (time is the vertical axis and increases from top to bottom in the image). Bars in the image with smaller angles with respect to the top of the image have larger speeds. Simultaneous increase of speed and size is evident because wider bars (larger random dots) are moving at larger speeds than more narrow bars[6].

The input to the simulation was generated by applying the filter in figure 6.7 to the motion input using a standard convolution. The resulting filter response was then half-wave rectified in accordance with the idea that only positive cell responses to visual stimuli are transmitted to the next layer of processing. The images were assumed to be symmetric for the convolution. That is, filter and motion image coordinates were wrapped in accordance with mapping the images onto a torus. This obviated handling of kernel-image edge effects. Figure 6.10 shows example images from this process. Each image represents the output of a bank of kernels that span the motion image and corresponds to the output of the LGN filter to the motion over time. Again, time is the vertical axis and increases from top to bottom.

The four images in figure 6.10 represent the extreme spatial and temporal scales that have been tested to date. The two images on the left represent the output of filters with center-Gaussian standard deviations of $\sigma_c = 20$ pixels. The two images on the right represent the output of filters with center-Gaussian standard deviations of $\sigma_c = 2$ pixels. The temporal response scale of the top two filters was a factor of four larger (longer) than the bottom two filters. The images show the non-lagged filter outputs. The lagged filter outputs are essentially the same with a shift of the output images on the time axis. Both the motion and filter kernel images were 512 pixels square.

[6] Note that there are some bars in the image that look very narrow in the space-time domain but are very "fast". Consideration of a single time slice will convince the reader that the pixel sizes for those bars is actually very large, and that the narrowness of the bars is due to the high speed of movement of those pixels.

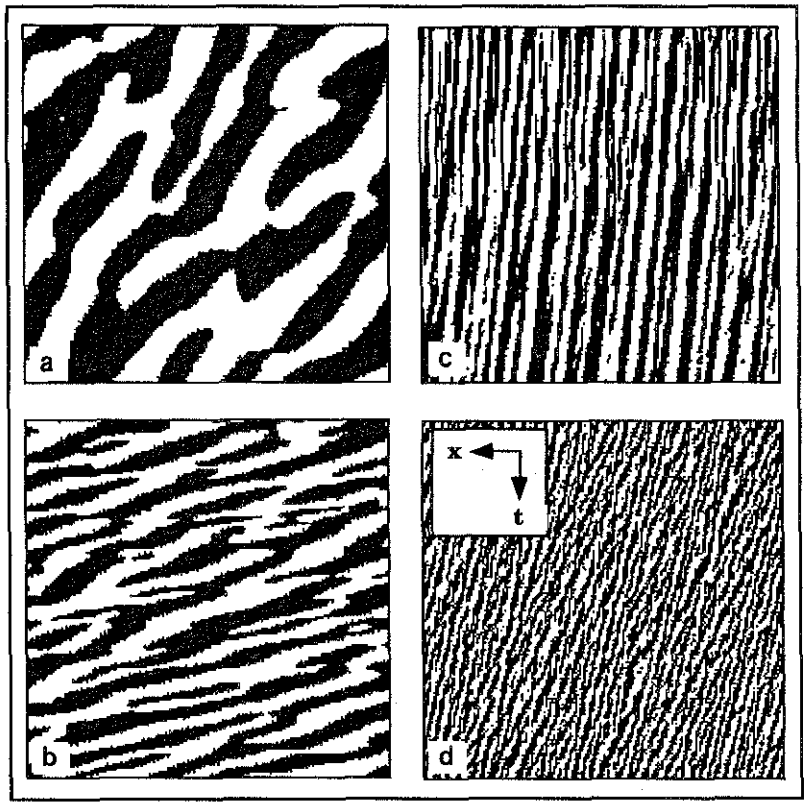


Figure 6.10.
The images represent LGN-like responses to the motion image in figure 6.9. Positive response magnitudes are represented by white pixels and zero response magnitudes are represented by black pixels in the image. Each column within an image represents the output of a filter at the corresponding spatial position in the motion image and corresponds to the output of the LGN filter to the motion over time (time increases from top to bottom). The spatial scale of the filters decreases by a factor of 10 from left to right. The temporal scale decreases by a factor of four from top to bottom.

It is clear from the images in figure 6.10 that the spatio-temporal filters respond to motion information that is a function of their combined spatio-temporal characteristics. Larger scale spatial receptive fields respond to larger scale objects in the image. Faster (smaller) temporal scales respond to faster velocities in the image. Relating the selected motion and filter scales to the real world would be extremely difficult, requiring specification of ranges of object sizes and distances (from the observer) in the real world. We would also have to specify the retinal image sizes and the corresponding LGN spatial receptive field sizes. Those steps are beyond the scope of this research. Input to the simulation was a subset of the filter image columns (positions of the filters across the motion image). The filter positions were chosen so as to span the motion image and to avoid wrapping the filter kernel to the opposite side of the motion image.

6.4.2.2 Architecture and details

The simulation architecture was an EXIN (excitatory-inhibitory) network as described by Marshall (1993). The network had one output and two input layers as shown in figure 6.11. The output layer consisted of 22 neurons. Each input layer consisted of 21 neurons, one for lagged-type filter input and one for non-lagged-type filter input. The lagged and non-lagged filter types were used to look for the emergence of directional selectivity that may result from having both lagged and non-lagged types of input. The effects of FT vs. SS response types were not considered because using various temporal response scales for one type (say FT) creates a range of temporal response characteristics that may overlap with the other type. That is, a very large temporal scale FT type filter kernel may be very similar to a SS type filter with some other temporal scale. Because of the possibility of confounding temporal scale with response type, only the FT type of filter was used in the simulations reported here. A detailed specification of the filter impulse responses used in the simulations is given in the appendix.

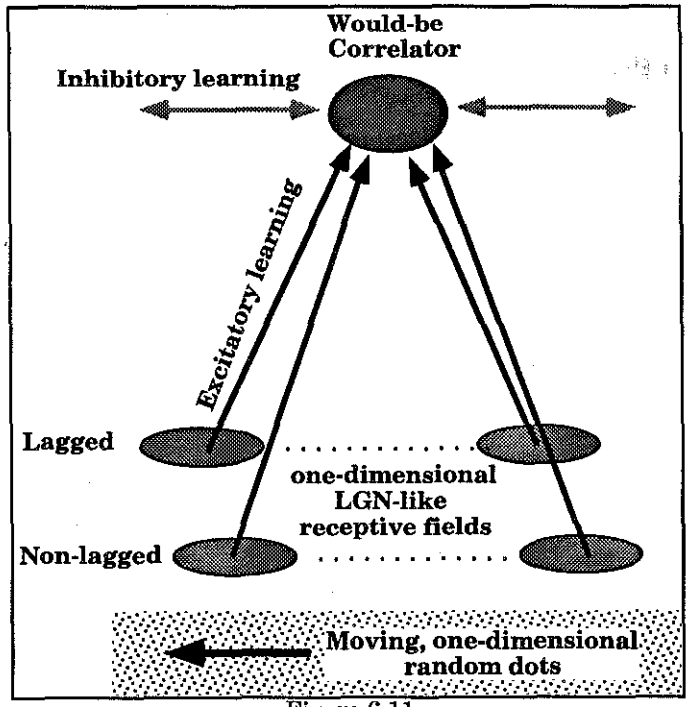


Figure 6.11
Network architecture of the computational simulation.

The input layers were fully connected to the output (pseudo-correlator) layer. The feed-forward connections were modified using an excitatory learning rule. The output layer employed lateral inhibitory connections within the output layer with an inhibitory learning rule. The lateral inhibitory connections were used to discourage duplicated input pattern selection by the output neurons. The inhibitory connections were modified using a learning rule designed to allow output neurons to select overlapping input patterns. In combination, the excitatory and inhibitory learning rules were designed to allow self-organization of networks in complex environments that include multiple superimposed patterns at multiple scales (Marshall, 1993). The learning rules and the output layer activation function are described in Marshall (1993) and reproduced in the appendix. The activation function of the output layer was designed to provide a response to a sufficient input magnitude rather than to mimic a multiplicative correlation operation (i.e., influence of input values was additive rather than multiplicative). The output of the input layers was the half-wave rectified response of the LGN-like filters to a composite motion input in which all motion was unidirectional[7]. The network connection weights were initialized to random values (see appendix). The network was then trained on the filter output signals (presented in chronological order) until the connection weights converged (learning stabilized). Convergence required up to 18,000 pattern presentations.

6.4.3 Results

The network attributes of interest here are the feed-forward connections to each output neuron and the relationship between the sets of connections to the lagged and non-lagged input subsets. Figure 6.12 shows connection weight data for typical neurons from three different simulations (i.e., after training). The vertical axis on each graph represents the connection weight, and the horizontal axis is the input neuron number. The input number specifies the relative spatial position of the input filter across the composite motion image. Each of the three graphs shows two curves: one for the connections to the lagged input neurons (no markers) and one for the connections to the non-lagged input neurons (black circle markers).

Lagged and non-lagged filters with the same input number had identical positions in the motion image. For the range of simulation parameters used so far, the non-lagged weight curves generally have one very strong peak with weaker peaks to one or both sides. The peaks in the lagged-weight curve are generally interleaved with the peaks of the non-lagged curve as shown in the figure. Across simulations the width of the peak appears to be proportional to the receptive field scale: large receptive fields have wider peaks in both the lagged and non-lagged connection weight curves.

[7] During the learning phase, the pattern motions were presented in chronological fashion. The coherent motion lasted 512 time steps (the size of the image) and then was repeated from the begining. The output of the filter therefore includes some motion noise that results from restarting the spatial pattern motion.

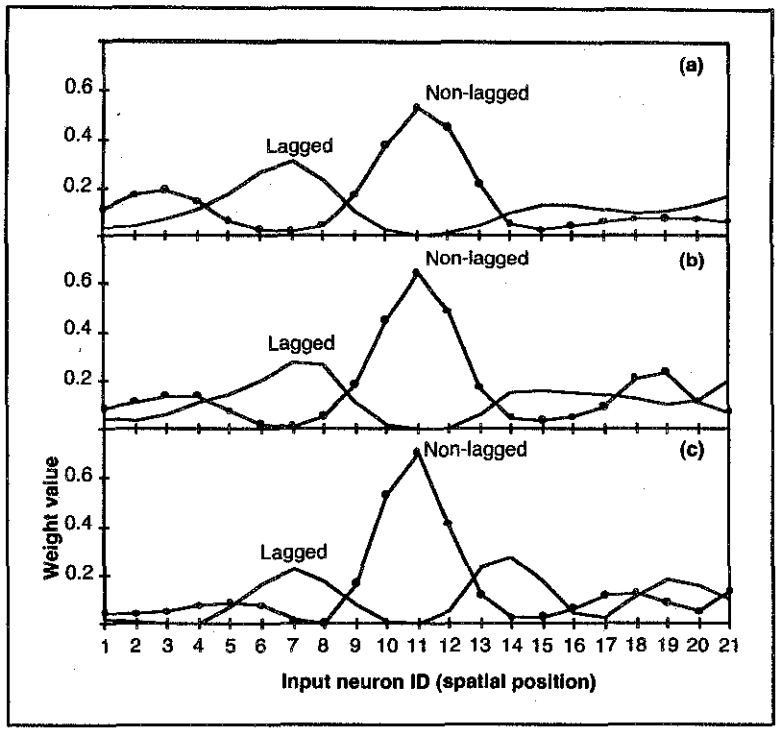


Figure 6.12.
Connection weight data for typical neurons from three different simulations. The vertical axis on each graph represents the connection weight and the horizontal axis is the input neuron number (relative spatial position). Each of the three graphs shows two curves: one for the connections to the lagged input neurons (no markers) and one for the connections to the non-lagged input neurons (black circle markers). The temporal response scale of the filter used in (b) is a factor of two longer than that used in (a). The temporal response scale of the filter used in (c) is a factor of 4 longer than that used in (a).

The different simulations shown in different graphs in figure 6.12 used very different temporal response scales. The temporal response scale increased fourfold from 6.12a to 6.12c but did not significantly influence the correlation distances chosen in each simulation. The three simulations used the same spatial receptive field size: a receptive-field-center standard deviation (σ_c) of 20 pixels with a receptive-field-surround standard deviation (σ_s) of 34 pixels. The filter spacing in the original motion input image was 14 pixels, and the distance between the largest lagged and non-lagged peaks was from three to four filter positions.

6.4.4 Discussion

The simulation output needs to be interpreted in order to answer a few interesting questions. We need to understand whether the connection weight data would implement a preferred correlation distance (an optimum spatial pattern displacement size) if the learned connection weights were used to specify LGN-like filter input strengths for a correlator neuron. We also need to understand whether a correlator neuron implemented using the connection weight data would result in speed selectivity and/or directional selectivity.

6.4.4.1 Correlation distance

We can consider the distance between the peak in the lagged curve and the peak in the non-lagged curve to correspond to the preferred spatial correlation distance ($\mathbf{S_d}$) for the would-be motion detector. The spatial correlation distance in all the panels in figure 6.12 is then 1.2 to 1.6 times σ_s. The input receptive fields selected by the neuron shown in figure 6.12 therefore overlap in visual space. Simulations with smaller receptive field sizes show the same result (not shown): an $\mathbf{S_d}$ that is in proportion to the spatial receptive field size with the same spacing factor (1.2 to 1.6 times σ_s). Figure 6.13 shows what this means in terms of the lagged and non-lagged spatial receptive fields. The figure shows the spatial profile of two receptive fields spaced by a factor of 1.5 times σ_s. (That is, one receptive field center is 1.5 x σ_s away from the other receptive field.) One possible consequence of this configuration is that a spatial pattern that optimally stimulates one of the receptive fields is likely to optimally suppress the other receptive field (assuming a somewhat regular spatial pattern).

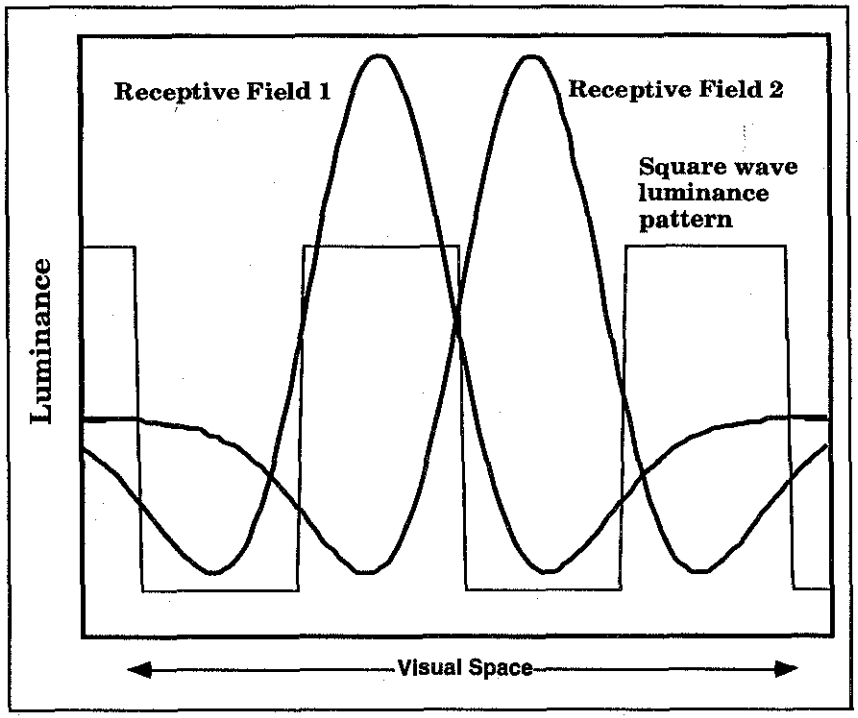


Figure 6.13.
Two spatial receptive field profiles are shown. The centers are offset in space by 1.5 times the standard deviation of the surround (negative) Gaussian used to create the DOG filter. An example square wave pattern has been added to illustrate that the distance between receptive field centers implies that an optimally activating spatial pattern for one of the receptive fields would generally be an optimally suppressing spatial pattern for the other receptive field.

6.4.4.2 Speed selectivity

The inputs with larger weights contribute more to the activation of the output neuron. In terms of a temporal activation pattern for the input neurons, the output neuron would respond best to simultaneous input from the filters that correspond to the peak positions on the weight curves. The delay inherent in the lagged filter input is therefore important. An optimal spatio-temporal input for the weight values in figure 6.12a would require an optimal spatial pattern to first activate the lagged filters around position 7 (a position in visual space), and then move to position 11 (a different position in visual space) within a time frame commensurate with the inherent delay of the lagged filter.

6.4.4.3 Direction selectivity

Some of the connection weight matrices also show what might be interpreted as representing direction selectivity (e.g. see figure 6.12a): an asymmetric arrangement of lagged connection-weight-curve peaks. The peak arrangement in figure 6.12a should respond to movement of an optimal pattern from position 7 to either position 11 or position 3. Between those two possible situations, the larger peak weight values in position 11 indicate that movement from position 7 to position 11 should result in a larger response magnitude of the output neuron. The result is that the output neuron would respond better to a direction of motion from position 7 toward position 11 than it would to a motion from position 7 toward position 3. This asymmetry is most prominent for the shortest temporal scale (figure 6.12a) and least prominent for longest temporal scale (figure 6.12c). The spatial and temporal scale of the receptive field in figure 6.12 was not well matched to any of the coherent motion present at the input (where a pixel's speed was proportional to its scale). The spatio-temporal filter output that produced the results in figure 6.12c was therefore a result of aliasing caused by that mismatch. The aliasing reduces or removes the temporal correlation between the lagged and non-lagged filters. The directional symmetry (lack of direction selectivity) is therefore consistent with the output of the filters. That is, formation of directional selectivity in a motion detector should require the availability of coherent motion during the formation of the detector.

6.4.4.4 Motion detector formation

In the simulation parameters tested to date, it appears that the spatial receptive field size determines $\mathbf{S_d}$ independent of the temporal response characteristics of the receptive field. However, the spatial and temporal response scales of the receptive fields *together* determine the preferred speed of the self-organized motion detector (e.g. see filter outputs in figure 6.10). This result was at first surprising; we might expect that $\mathbf{S_d}$ would increase with preferred speed. In retrospect, however, the simulation results are consistent with the critical width behavior seen in figures 5.6 and 5.9 for subject EF. The critical width (cf. figures 5.6 and 5.9, subject EF) appears to be correlated with the spatial receptive field size but *not* the

stimulus frame rate (motion speed), at least with regard to the ranges of parameters tested so far in the psychophysical experiments. That is, the speed of the motion in the stimulus did not affect the critical dimension sizes.

To provide a possible explanation of the self-organized detector characteristics represented by the connection weight curves in figure 6.12, we must consider the simulation learning rules, the properties of the input signal, and the properties of the LGN-like filters. Perhaps most importantly, the learning rules are designed to provide strong connections from the input (i.e., the LGN-like filter outputs) to the would-be correlator neuron. The learning rules are designed to make connections to filter (input layer) outputs that are most strongly correlated with the would-be correlator neuron response. Thus we should interpret the response of filters in terms of correlation between filter outputs over "visual space" and stimulus duration.

Correlating filter outputs

Because the filter kernel used here was separable in space and time, the temporal correlation of the filter outputs can be considered separately. The crosscorrelation over space of the LGN-like spatial filter responses provides an indication of the spatial correlation distances that might be possible. Because the two filter types had the same spatial filter kernel we can just calculate the autocorrelation of that kernel. The input to the filter array is spatially uncorrelated random dots. Call the output signal **y** and the autocorrelation function $\phi_{yy}(\tau)$ where τ is the correlation distance. Then $\phi_{yy}(\tau)$ can be calculated from the spatial power spectral density (PSD) of the input signal and the spatial frequency response of the input filter (Proakis, pp. 50-52, 1983). The input signal (call it **x**) is uncorrelated random noise, so the spatial PSD of the input signal is constant over all spatial frequencies ($\Phi_{xx}(f) = N_0$). Call the frequency response of the spatial DOG part of the LGN-filter **H**(f). The PSD of the filter output is defined in equation (6.11).

$$\Phi_{yy}(f) = N_0 \cdot |H(f)|^2 \tag{6.11}$$

$\phi_{yy}(\tau)$ is just the inverse Fourier transform of $\Phi_{yy}(f)$. We can take a shortcut to calculate the autocorrelation because we know that multiplication in the frequency domain is equivalent to convolution in the spatial domain. Thus our spatial autocorrelation function for the filter output is the convolution of the filter's spatial impulse response with itself, all multiplied by N_0. Figure 6.14 shows an example comparison of the spatial impulse response of the LGN-like filters together with the autocorrelation function for that filter (the first is a function of space, the second is a function of offset).

All non-zero parts of the spatial impulse response and autocorrelation functions are shown in the figure. For random dot pattern inputs, it is clear that spatial correlations only exist across a limited spatial distance determined by the filter size. This is consistent with a self-organized $\mathbf{S_d}$ that is a function *only of the spatial scale of the filter* (commensurate with the simulation results). Moreover, the autocorrelation function is symmetric about the center of the filter position. We should therefore expect that in the presence of motion noise input (uncorrelated in time) direction selectivity would not appear in the simulation because motion noise contains information in all directions.

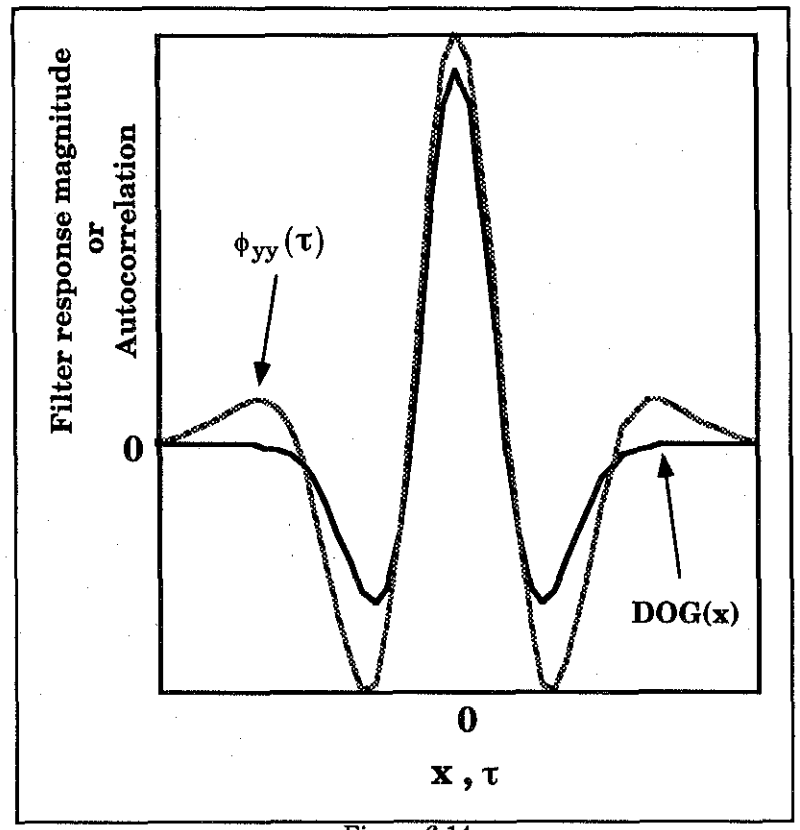


Figure 6.14
The DOG(x) impulse response of the LGN-like input filters compared to its autocorrelation function.

Optimal offset distances

The learning rules are designed to select inputs that are well correlated in time. The temporal response of the filter will determine the temporal correlation of the filter outputs by modifying $\phi_{yy}(\tau)$ multiplicatively (because the filter kernel is separable). In this simulation that modification includes a reversal of sign (negative vs. positive correlation) for both the lagged and non-lagged filters, so there is some point in the spatio-temporal temporal correlation at which the negative lobes of $\phi_{yy}(\tau)$ become positive. Those offset positions must therefore be included in our consideration of possible lagged vs. non-lagged offset distances.

According to figure 6.14, then, there are three spatial offsets ($\mathbf{S_d}$ values) that are optimal for a lagged and non-lagged filter pair. The position with the largest correlation magnitude a zero offset, or complete overlap of the two filters. The next best offset is as shown in figure 6.13, and the least best offset is the center of the smallest peak for $\phi_{yy}(\tau)$ near the edge of figure 6.14. The initial expectation, given that the learning rules create connections to select the greatest amount of correlation, is that a correlator neurons should always select lagged and non-lagged filter pairs that overlap because that provides the best spatial correlation. However, because of the lag included in the lagged-filter response, it is physically impossible to get temporal correlation in that situation (assuming that the lag is long enough). In such a situation, both the lagged and non-lagged filters are presented with the same spatial pattern and respond to that pattern identically, except that the lagged filter response to the pattern occurs some time later. The output of the lagged filter at the time of a new pattern presentation represents some spatial pattern that was shown *some time in the past* (determined by the lag). The only possible way to get correlated outputs for the filter pair in overlapping case, then, is to present a static or very slowly moving pattern to the two filters. This type of motion was not included in the input to the simulation. Moreover, in the simulations presented here the filter responses are not of a sustained type, so even the static pattern should fail to cause development of strong connections to the same output neuron for a lagged and non-lagged filter pair in the same position. Elimination of the overlapping case (because it is not possible to implement) leaves us with the "second best" case, which happens to be the configuration that the learning rules developed for us.

Development of asymmetry vs. symmetry

The next question is how coherent motion might cause an asymmetric connection-weight pattern (direction selectivity) between the lagged and non-lagged filters. Coherent motion in a single direction should bias the amount of correlation between the lagged filters to either side of a given non-lagged filter. For example, let us consider the connection-weight scheme in figure 6.12a. The lagged filter in position 7 is correlated with the non-lagged filter in position 11 if an appropriate spatial pattern

appears first at position 7 and then appears at position 11 after a time commensurate with temporal response properties of the filters (and especially the lag of the lagged filter). However, the very presence of coherent motion in a single direction precludes good correlation between the non-lagged filter at position 11 and the alternative filters around position 15 *because of the built-in lag of the lagged filters*. In this case the spatial pattern is moving in a direction *opposite* to that which would result in a good correlation of filter outputs over time.

The use of unidirectional motion input causes a significant bias in the temporal correlation between the lagged and non-lagged filter outputs. The temporal correlation therefore biases the symmetric spatial correlation that we begin with (see figure 6.14), but the temporal correlation does not cancel out random correlations in the wrong direction (motion noise, or false correlations). The excitatory learning rules used in the simulation are designed to select the input patterns with the *greatest degree* of correlation. The bias in correlated *spatio-temporal outputs* between the lagged and non-lagged filters that results from unidirectional motion input therefore causes an asymmetric connection weight pattern.

The lack of directional selectivity (the symmetric lagged peaks) in figure 6.12c is therefore not surprising. The filters had a large spatial scale and a large temporal scale (a low preferred speed). The large spatial scale dot patterns in the composite motion input *all moved at high speed*. The response of the large scale filters tuned to low speeds is a result of aliasing caused by the filter-motion mismatch, and that aliasing eliminates any temporal correlation bias for one of the possible filter offset directions. The connection weight curves in figure 6.12c are therefore symmetric because there was no directional motion in the composite motion image that could drive the formation of directional selectivity.

6.5 General Discussion

In order to create a motion detector, the would-be correlator neuron must select filters with separated positions in visual space. Selection of both lagged and non-lagged filters allows the extraction of motion information from the "visual environment" in the simulation. The analysis in section 6.4.4 indicates that, for the simulation parameters tested here, spatial correlation distances might be predicted by the autocorrelation of the filter output across space. The spatial autocorrelation is independent of the temporal response properties of the filters. Additional temporal correlation restrictions resulting from the use of both lagged and non-lagged filter types (i.e., the presence of a lag) may allow the emergence of direction selectivity when the input contains coherent, or unidirectional motion.

The relation of these results to the actual biological system are not conclusive regardless of the care taken in designing the simulation because the simulation is not equivalent to the biological organism. However, the results are consistent with discussions of the possible organization of the motion detector array in the previous chapters. Moreover, the independence of $\mathbf{S_d}$ from the temporal response of the detector is consistent with the psychophysical data in chapter 5 (as mentioned above). If these results apply to the organization of biological systems, some of the next questions to be asked might be (1) What is the range of available spatial and temporal receptive field scales when the motion detector array is created? (2) How are spatial and temporal scales related in the physiological system? (3) What combination of spatial and temporal receptive field properties exist for use in the motion detection system? If the answer to the these questions is that there enough to cover the available motion information, we might look to equation (6.10) for answers. At any angle Θ relative to the point of gaze, there is a range of visible objects (the range of visible spatial scales) at a range of distances $\mathbf{d}$. The images of these objects are moving on the retina and represent the range of visual motion speeds available to the motion detection system at that position in the visual field. As Θ changes, that range of speeds changes accordingly.

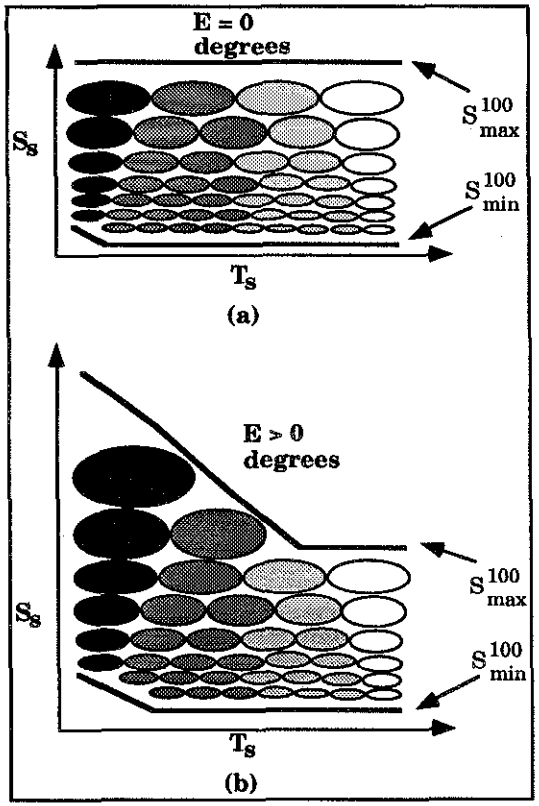


Figure 6.15.
(a) A layout for motion detector receptive fields in the foveal region of visual space, and (b) a layout for some position in the visual field outside the fovea. The receptive fields are shown as ellipse icons mapped onto the $<S_S,T_S>$ plane. The position of each ellipse represents the $<S_S,T_S>$ values to which the detector is tuned. Increasing size indicates increasing spatial scale of the detector's composite receptive field. Dark to light shading indicates increasing scale of the detectors temporal response. That is, dark represents a fast response while light represents a slower response.

In light of the psychophysical and simulation results we can propose that the distribution of receptive field spatial and temporal scales of the motion detector array may follow **E** as shown in figure 6.15 (assuming that motion detector receptive field characteristics are correlated with the characteristics of their input receptive fields). Figure 6.15a shows a possible motion detector receptive field layout for the foveal region of visual space, and figure 6.15b shows a possible layout for some non-zero **E** value. The receptive fields are shown as ellipses (icons) on the $<S_S,T_S>$ plane. The increasing ellipse size indicates increasing spatial scale of the detector's receptive field. Dark to light shading indicates increasing temporal scale for the detector's receptive field response.

The position of an ellipse indicates that a detector of the spatio-temporal response properties indicated by the size and shading of the ellipse prefers that $<S_S,T_S>$ value in an apparent motion stimulus. The sizes and shading of the icons in the figure are not intended to indicate any absolute values. Instead, the icons indicate a relative distribution of the spatial and temporal response properties of the motion detectors that serve a given region in the visual field (as a function of **E**). The smallest and largest visible displacement sizes, called S^{100}_{min} and S^{100}_{max} in the previous chapters and in the figure, are represented by the dark lines that demarcate the region of the $<S_S,T_S>$ plane covered by the available detector array. The shapes for S^{100}_{min} and S^{100}_{max} in the figure correspond to those in figure 6.1.

The key concept here is that the available spatial and temporal response characteristics may change with **E**, but the relative distribution of the properties on the $<S_S,T_S>$ plane follows the same pattern in each situation. The distribution should be determined by self-organization of the motion detection system driven by the motion information presented at a given portion of the visual field. Preferred spatial displacement scale is proportional to spatial receptive field size while preferred speed is determined by the temporal response scale. Each spatial receptive field scale has a range of possible temporal response scales. Figure 6.12b indicates that that range decreases as spatial scale increases for regions in the visual field away from the fovea.

It may be possible to test the hypothesized distribution of spatial and temporal scales in the $<S_S,T_S>$ plane by using spatially filtered versions of the visual motion stimulus used in the experiments presented in chapters 2-5. A spatially filtered image could provide improved isolation

of receptive field sizes. Furthermore, if this layout is driven by the available motion information, the cosine component of equation 6.10 should be visible in the scaling of S_{min}^{100} and S_{max}^{100}. That is, the scaling of S_{min}^{100} and S_{max}^{100} should only be linear for small angular distances from the fovea.

6.6 Conclusions

The psychophysical and simulation results appear to be consistent when viewed from the point of view of equation (6.9) and figure 6.15. However the combination of the equation and figure raise at least as many questions as have been answered by the results. Moreover, there are many other combinations of simulation parameters that should be tested. For example, the influence of the FT vs. SS type of temporal response should be tested, as well as combinations of on-center and off-center receptive field types. The predictive nature of equation (6.9) and figure 6.15 provides the basis for designing further psychophysical and simulation experiments that can verify or falsify the overall model.

Acknowledgments

R. E. Fredericksen was supported by a an International Brain Research Organization Research (IBRO) Fellowship. I thank W. A. van de Grind and F. A. J. Verstraten for valuable discussions of the model formulation. Special thanks go to J. A. Marshall for providing access to his neural network simulation software and for discussions on the structure of the simulation. Final thanks go to Richard Alley for modifications to the simulation software that allowed me to finish the simulations in time for this dissertation.

Bibliography

Baker, C.L. Jr. and Braddick, O.J. (1985) Eccentricity-dependent scaling of the limits for short-range apparent motion perception. *Vision Research*, **25**, 803-812.

Drasdo, N., (1991) Neural substrates and threshold gradients of peripheral vision. In Vision and Visual Dysfunction, Vol. 5, Macmillan Press, pp. 251-265.

Emerson, R.C., Bergen, J.R. and Adelsen, E.H. (1992) Directionally selective complex cells and the computation of motion energy in cat visual cortex. *Vision Research*, **32**, 203-218.

Fleet, D.J. and Jepson, A.D. (1985) Spatiotemporal inseparability in early vision: centre-surround models and velocity selectivity. In Computational Intelligence, Volume I, pp. 89-102.

Luhmann, H.J., Martínez Millán, L., and Singer, W. (1986) Development of horizontal intrinsic connections in cat striate cortex. *Experimental Brain Research*, **63**, 443-448.

Marshall, J.A. (submitted) Adaptive perceptual pattern recognition by self-organizing neural networks: Context, uncertainty, multiplicity and scale.

Mastronarde, D.N. (1987) Two classes of single-input X-cells in cat lateral geniculate nucleus. I. Receptive-field properties and classification of cells. *Journal of Neurophysiology*, **57:2**, pp. 357-380.

Mastronarde, D.N. (1987) Two classes of single-input X-cells in cat lateral geniculate nucleus. I. Retinal inputs and the generation of receptive-field properties. *Journal of Neurophysiology*, **57:2**, pp. 381-413.

Proakis, J.G., Digital Communications, McGraw Hill Book Co., 1983.

Reid, R.C., Soodak, R.E. and Shapley, R.M. (1991) Directional selectivity and spatiotemporal structure of receptive fields of simple cells in cat striate cortex. *Journal of Neurophysiology*, **66:2**, pp. 505-529.

Saul, A.B. and Humphrey, A.L. (1990) Spatial and temporal response properties of lagged and non-lagged cells in cat lateral geniculate nucleus. *Journal of Neurophysiology*, **64:1**, pp. 206-224.

Wattam-Bell, J. (1991) Development of motion-specific cortical responses in infancy. *Vision Research*, **31**, 287-297.

Wattam-Bell, J. (1992) Development of maximum displacement limits for discrimination of motion direction in infancy. *Vision Research*, **32**, 621-630.

Appendix A

Motion Image

A single motion image was used in all the simulations. The motion image was a composite of several one-dimensional random dot patterns at different spatial scales and velocities. The motion image was 512 pixels square. Pattern velocities were produced by displacing the random dot patterns by **dx** pixels every **dt** time steps. Random dots in the image were integer multiples (**m**) of the pixel size. The combinations of dot sizes and velocities (**m, dx, dt**) were (1,1,8), (2,1,4), (4,1,2), (8,1,1), (16,2,1), (32,4,1), and (64,8,1). The random dots were chosen to be black or white with equal probability. The absolute values for the black and white dots in each motion image were symmetric about zero and were the same for each individual pattern. Each motion image represented by the above (**m, dx, dt**) combinations was created separately and then added together (luminance addition) to form the composite motion image in figure 6.9. The pixels in the composite image were scaled to the range [0,255] and had a mean value of 128.

LGN-like Filters

The LGN-like filters were 512 pixels square. The filters were created using a difference-of-Gaussian (DOG) approach in both space and time. The spatial and temporal responses were separable, so the overall filter response was defined as

$$\text{(A1)} \quad DOG(x,t) = DOG(x) \cdot DOG(t)$$

The spatial part of the filter was created by a DOG in space that approximated a Laplacian of a Gaussian. This was achieved by setting the standard deviation of the negative spatial Gaussian to be 1.7 times that of the positive spatial Gaussian. We can designate the spatial part of the DOG filter by the standard deviations of the center (σ_c) and surround (σ_s). Thus a DOG(x) filter with $\sigma_c = 20$ also had $\sigma_s = 34$. The spatial scale of the filter can be designated simply by σ_c.

The temporal responses of the lagged and non-lagged filters were also created using the DOG approach, except that the means of the two

Gaussians were offset. The values chosen were selected by matching the temporal impulse responses derived from physiological data by Saul and Humphrey (1990) (cf. their figure 11). Each filter can be designated by a mean and standard deviation for the two Gaussians. Let the mean and standard deviation of the positive Gaussian be μ^+ and σ^+, respectively. Likewise, let the mean and standard deviation of the negative Gaussian be μ^- and σ^-, respectively. The largest scale lagged and non-lagged filters were (in pixel units)

	μ^+	σ^+	μ^-	σ^-
Lagged	25	105	90	45
Non-lagged	20	35	30	85

Figure A.1

The parameter values for the largest scale filters used in the simulations.

Smaller temporal scale responses were achieved by successively dividing the above parameter values by a factor of 2. If we designate the filter described in the above figure as temporal scale 0 (TS_0), then TS_k for non-negative integer k are filters of smaller temporal scale defined by

$$\text{(A2)}\quad TS_k = \left(\left\lfloor \frac{\mu^+}{2^k} \right\rfloor, \left\lfloor \frac{\sigma^+}{2^k} \right\rfloor, \left\lfloor \frac{\mu^-}{2^k} \right\rfloor, \left\lfloor \frac{\sigma^-}{2^k} \right\rfloor \right)$$

The filters in figures 6.10a and 6.10c had temporal scale 0. The filters in figure 6.10b and 6.10d had temporal scale 2. The filters in figures 6.10a and 6.10b had spatial scale 20. The filters in figures 6.10c and 6.10d had spatial scale 2. The filters that produced the connection weight data shown in figure 6.12 were all of spatial scale 20. Figures 6.12a through 6.12c were temporal scales 2, 1, and 0, respectively.

Simulation Input

The output values of the input layers were produced by convolving the filter images with the composite motion image, thereby applying the LGN-like spatio-temporal filter to every spatial position in the composite motion image. The output of the convolution process was then half-wave rectified to remove "negative" neuronal responses. Filter positions for the 21 neurons in each input layer for each simulation were selected so that the filters overlapped in "visual space." The filter positions were selected so that they spanned the central portion of the composite motion image and were sufficiently distant from the image edge to avoid effects that result from wrapping the filter kernel around to the opposite side of the motion image. The sequence of input vectors to the network during learning was a (chronologically correct) cyclic presentation of the half-wave rectified filter output.

Learning and Activation Rules

Activation changes for layer 2 neurons were defined by

(A3)

$$\frac{d}{dt}x_i = -Ax_i + (B - x_i)\frac{\beta \sum_j \left(m(x_j) z_{ji}^+ \right)}{\alpha + \sum_j z_{ji}^+} - \gamma (C + x_i) \sum_j \left(m(x_j) z_{ji}^- \right)$$

where x_i represents the activation value of neuron i, z_{ji}^+ represents an excitatory connection from neuron j to neuron i, z_{ji}^- represents an inhibitory connection from neuron j to neuron i, $m(x_j) = \max(x_j, 0)$, and **A**, **B**, **C**, α , β , and γ are constants. Changes in the weights were determined by

$$\text{(A4)}\quad \frac{d}{dt}z_{ji}^+ = \varepsilon \cdot f(x_i)\left(-z_{ji}^+ + h(x_j)\right)$$

$$\text{(A5)}\quad \frac{d}{dt}z_{ji}^- = \delta \cdot g(x_j)\left(-z_{ji}^- + q(x_i)\right)$$

where ε and δ govern the learning rates, and with

$$
\begin{aligned}
f(x_i) &= m(x_i)^2 \\
h(x_j) &= H \cdot m(x_j) \\
g(x_j) &= m(x_j) \\
q(x_i) &= Q \cdot m(x_i)
\end{aligned}
\tag{A6}
$$

where **H** and **Q** are constants.

Network Structure

All simulations used the same network structure. There were two input layers of 21 neurons each. The two input layers had output values that were the lagged and non-lagged LGN-like receptive field responses to the motion input. The spatial and temporal scale of the LGN-like receptive fields was varied across simulations. The input layers were always fully connected to the output layer. The output layer inhibitory (intra-laminar) connections were always fully connected within the output layer.

Initial Connection Weights

Let R_{ji} be drawn pseudo-randomly from the interval [0,1). The initial excitatory connection weights from layer one to layer two were defined as

$$z_{ji}^{+}(0) = Z_0^{+}\left(1 + V^{+}\left(2R_{ji} - 1\right)\right) \tag{A7}$$

and the initial inhibitory connection weights were defined as

$$z_{ji}^{-}(0) = Z_0^{-}\left(1 + V^{-}\left(2R_{ji} - 1\right)\right) \tag{A8}$$

Simulation Parameters

$A = 20$	$B = 1.0$	$C = 0.1$
$\alpha = 1.0$	$\beta = 12.5$	$\gamma = 50000$
$\delta = 0.075$	$\varepsilon = 0.075$	
$H = 1$	$Q = 1$	
$Z_0^{+} = 1.0$	$Z_0^{-} = 0.25$	
$V^{+} = 0.01$	$V^{-} = 0.01$	

Computer Implementation

The simulations were run on a MasPar MP-1 parallel computer with 8192 processors. The simulation software used the Euler method to numerically integrate the differential equations.

Chapter 7: Summary and Conclusions

7.1 Introduction

The first six chapters have each presented investigations of various aspects of our ability to perceive motion in random dot apparent motion stimuli. Section 7.2 summarizes the results of each of the previous chapters. Section 7.3 then discusses how the results fit together into an integrated model. That model is based on experiments covering several different parameters in the stimulus, but the experimental results presented in the chapters has not exhaustively tested all seven axes. Section 7.4 discusses caveats and omissions for the model. Section 7.5 presents some experiments that support model extensions, and ideas on how the extensions might be made. Section 7.6 presents some ideas for extended research directions that are not direct verifications of the integrated model. Section 7.7 enumerates some of the motion perception phenomena that are not addressed by the integrated model. Finally, section 7.8 briefly discusses how the basic research presented in this dissertation may be useful in the computer science and engineering community.

7.2 Summary of results

Chapter 2 begins with the formulation of a novel experimental paradigm for use in the study of motion perception. The stimulus and threshold technique by themselves are not new, but their use in combination is. The paradigm uses a visual motion stimulus designed to isolate motion detectors that are tuned to a specific $\langle S_S, T_S \rangle$ combination (single lifetime random dot patterns) together with a method for measuring the visibility of the stimulus (luminance signal-to-noise ratio). The result is an experimental method that allows the measurement of the visibility of motion stimuli containing coherent motion at only a single combination of spatial displacement size and frame rate. The method was used to measure threshold surfaces over the $\langle S_S, T_S \rangle$ plane.

The experimental results in chapter 2 refine and extend previous results on the visibility of large and small S_S values, and extend visibility measurements to include the entire range of $\langle S_S, T_S \rangle$ values. The differences in previous results and the results reported here can be attributed to the use of a stimulus duration that is long enough to allow a maximum response of the motion detection system (due to temporal integration). An architectural implementation of the motion detection system is proposed. The architecture is based on known neuroanatomy and accepted cortical processing concepts: the mapping of visual space to visual cortex followed by a constant distribution of cortical interconnection distances that mediate the correlation of information by the motion detection system. A mathematical description of that architecture is given and is shown to be consistent with the differential dependence of S_{max}^{100} and S_{min}^{100} on E. The proposed motion system architecture provides a unified explanation of previous research regarding the visibility of $\langle S_S, T_S \rangle$ combinations. Moreover, the architecture provides a basis for explaining the visibility of $\langle S_S, T_S \rangle$ combinations *between* the visible extremes defined by S_{max}^{100} and S_{min}^{100}.

Chapters 3-5 investigate the spatial and temporal filtering characteristics of motion perception. Chapter 3 tests the improvement of perceptual thresholds with the stimulus duration (D_S) as a function of the number of displacements (N_D). The data agree with previous experiments with regard to improvement of S_{max}^{100} with N_D but show that, contrary to previous results, S_{min}^{100} can be improved in a similar manner. Furthermore, the data indicate that D_S is a better predictor of threshold improvement than N_D, and that temporal tuning of motion detectors is inversely related to the size of the spatial pattern displacement. In combination with the data in chapter 2 the data in chapter 3 show that, contrary to conclusions by some researchers, the range of visible $\langle S_S, T_S \rangle$ combinations can be accounted for by a single motion detection mechanism (rather than one for small S_S and another for large S_S).

Perhaps most importantly, the finding that D_S is a more precise indicator of threshold improvement than N_D provides the basis for designing and executing the experiments in chapter 4 which explore the interaction of T_S (the inverse of frame rate) and N_D as a function of constant D_S.

Chapter 4 is a continuation of the investigation of temporal filtering in the motion system using the more appropriate measure of total stimulus duration defined as $D_S = T_S \cdot (N_D + 1)$. The chapter presents a model for the temporal integration of random dot apparent motion that is constructed from neurophysiological data as well as psychophysical threshold measurements. The mechanism is based on the leaky integration of pulsatile motion detector responses to apparent motion stimuli. The pulses correspond to the response of motion detectors to discrete spatial displacements in the apparent motion stimulus. The psychophysical results indicate that temporal (luminance) contrast sensitivity determines the shape of constant-D_S threshold curves for image frame rates greater than about 7.5 Hz. Threshold values for frame rates less than 7.5 Hz are determined by the leaky integrator time constant and the shape of the pulses emitted by the motion detectors. The time required for the leaky integrator output to stop improving (response saturation) depends on the spacing of the output pulses (and therefore T_S). A similar dependence appears in the psychophysical data. The psychophysical data also indicate the influence of an absolute perceptual threshold for the populational response of the motion detector array. This absolute threshold is included in the temporal integration model. The leaky integration model provides a single, physiologically based explanation of temporal integration that allows unification of a range of previous experimental results.

Chapter 5 presents evidence from investigations of the filtering of motion detectors over visual space. The results show that spatial filtering follows the form of probabilistic summation of detector outputs for a wide range of S_S values. The probabilistic summation law holds when changing either motion stimulus height or width, a result that is different than those obtained using unlimited lifetime dot motion. Probabilistic summation breaks down when the motion display region approaches a particular height and/or width. These values, the critical height and width, increase with spatial displacement size and can be interpreted as a measure of the dimensions of the motion detector's receptive field. The critical height is smaller than the critical width, a result that is consistent with motion detector receptive fields that are elongated in the direction of motion. In addition, the mechanisms of temporal integration and spatial summation are shown to work independently under the conditions tested here. The results in chapter 5 refine and confirm previous results concerning the spatial summation of motion information. More importantly, the results presented in chapter 5 show that the spatial summation and temporal integration mechanisms can operate independently.

Chapter 6 uses the spatial summation result presented in chapter 5 to further expand the motion-detection system architecture proposed in chapter 2. A distribution of cortical interconnection distances ($P(D_c)$, where D_c is a distance in cortex) is fitted to psychophysical data using the equations derived and presented in chapter 2. The resulting distribution of D_c values is within the range that might be expected in primate or human cortex. One aspect of the psychophysical data that is not explained by a constant architectural model is the changing T_S-dependence of S_{max}^{100} and S_{min}^{100} with eccentricity E. Chapter 2 proposes that the change in dependence is a result of developmental forces that influence the creation of motion detectors covering each portion of the visual field. The detectors are likely to be created from a pool of front-end receptive field components that are possibly located at the level of the lateral geniculate nucleus in primates. Likewise, the correlator neurons are likely to correspond to a population of neurons in visual cortex that correlate front-end receptive field outputs over visual space. A computational simulation is presented that is designed to explore the influence of the spatial and temporal properties of the front-end receptive fields on the formation of correlation distances ($P(D_c)$) in visual cortex. The simulation results indicate that preferred spatial correlation distance (S_d) in the self-organized detector is proportional to front-end receptive field size and that the modeled temporal response properties of the front-end receptive fields do not influence S_d but do determine the preferred visual velocity of the detector. The elongated motion detector

receptive fields indicated by the critical dimension results in chapter 5 are therefore consistent with the simulation results. The computational simulation results together with the critical dimension results from chapter 5 allow the formulation of hypotheses about the relationship between the spatial and temporal properties of a motion detector and that detector's preferred motion velocity.

7.3 An integrated model

The stimulus parameters tested in the experiments presented here were

(1) spatial displacement size, $\mathbf{S_S}$,
(2) frame duration, $\mathbf{T_S}$,
(3) spatial displacement count, $\mathbf{N_D}$,
(4) total stimulus duration, $\mathbf{D_S}$ (derivable from $\mathbf{T_S}$ and $\mathbf{N_D}$),
(5) the position of the motion stimulus center in the visual field, $\mathbf{E_C}$,
(6) the stimulus height, **H**,
(7) and the stimulus width, **W**.

The finding of independence of spatial and temporal filtering in chapter 5 (for some combinations of stimulus parameters) allows us to merge the experiment-specific models presented in the individual chapters into a single, integrated model. That integrated model has interpolative and predictive power for the combinations of these seven parameters that have not been tested. The interpolative and predictive nature of the model allows for future psychophysical and neurophysiological testing that can verify or falsify the integrated model. We can write the model as

$$(7.1)\quad \mathrm{LSNR} = \frac{1}{g(L,\Psi,\cdots)\cdot\left(\frac{i(D_S)}{i(D_{S-min})}-1\right)\cdot\sqrt{N^*(S_S,T_S,E_C,H,W)}}$$

where the function g() includes the influence of stimulus luminance (**L**) and the spatial structure in the random dot patterns (Ψ), and the ellipsis indicates the presence of other possible stimulus parameters, such as color spectral content or competetive/cooperative interactions between detectors. The other parts of the denominator correspond to the temporal integration and spatial summation models described in chapters 4-6. The function i() is the leaky integrator response (chapter 4) and $D_{S\text{-}min}$ is the minimum stimulus duration required for visibility of the stimulus. In chapter 6, the function N() calculates the number of detectors under a one-dimensional model stimulus. The function $N^*()$ in equation (7.1) is the logical extension of the function N() evaluated by taking the additional step of integrating N() over the height of the (2-dimensional) stimulus.

7.4 Integrated model caveats

The proposed integrated model is interesting but requires the voicing of a number of caveats. The caveats stem from the lack of exhaustive testing of all stimulus parameter combinations. For example, the temporal integration and spatial summation properties were tested only in the central (2 to 8 degree square) portion of the visual field. The performance of the mechanisms may be different at other positions in the visual field. Furthermore, the temporal integration and spatial summation properties are shown as separable from each other and from the influence of overall stimulus contrast and the spatial structure in the random dot pattern. This assumption has yet to be verified. Moreover, evaluation of the function $N^*()$ can pose a problem. We can extend our N() in chapter 6 by assuming that $P(\mathbf{D_C})$ is homogeneous with respect to position in the visual field. We then multiply N() by **H** to get $N^*()$. However it is known that there are inhomogeneities and anisotropies in retinal architecture (retinal ganglion cell densities) and in psychophysical measurements of motion perception. Use of an $N^*()$ as described above might be a reasonable approximation but will not be completely accurate without incorporating measures of these position-dependent architectural/perceptual properties. Finally, the function N*() does not include the influence of the critical dimensions, i.e. the size of the motion detector receptive fields. The critical dimension limits on the size of the display should be included as a function of $\mathbf{S_S}$ in the evaluation of N*().

A further caveat comes from the use of a linear approximation to the cortical magnification factor (chapter 2; chapter 6). The linear approximation is only good to about 20 degrees away from the fovea. Corrections for some psychophysical and anatomical measures have been proposed that include other, non-linear terms in the equation. Such

changes will not affect the qualitative nature of the motion system architecture proposed here, but the changes can affect the correctness of the quantitative values predicted by the model.

A final caveat that must be considered is the use of numerical curve fitting for the cortical distribution $P(D_C)$ and the variation of E_2 with T_s (chapter 6), and the selection of pulse shapes in the temporal integration simulations (chapter 4). The goodness of the models must be considered in light of the use of those data-fitting methods. It would be preferable to use actual neurophysiological or neuroanatomical data, but in each of these cases the only available path was the curve fitting approach. Use of too many free parameters in such a data fitting process can be dangerous because a function can be fit to any data if the function has enough free parameters. However, in each case I used available psychophysical, neurophysiological, and neuroanatomical results and concepts to constrain the fitting process. Furthermore, the functions that were fitted contained a limited number of free parameters. It is therefore believable that the power of the experiment-specific models presented in the individual chapters and the integrated model presented above is not compromised by the data fitting approach.

7.5 Model extension/verification experiments

There are a number of model extensions that are possible with further psychophysical tests. Some of these extensions are described below.

7.5.1 Temporal integration and spatial summation outside the fovea

The first verification experiments that should be performed are testing of the temporal integration and spatial summation mechanisms outside the fovea. Unfortunately, application of the experiments presented in chapters 4 and 5 at non-zero eccentricities is not straightforward. Low contrast images in peripheral vision are prone to fading to invisibility when they are stationary in the peripheral visual field; this is called the Troxler effect. Good fixation by the subject can easily induce the Troxler effect in the motion stimulus presented here when the stimulus is away from the fovea, and more importantly, when the size of the dots (structures) in the image are below visual acuity or of low contrast. Pilot experiments have shown that extension of the temporal integration experiments outside the foveal region is possible, but care must be taken when scaling the stimulus. In chapter 2 the stimulus size was scaled according to previous measures of the scaling of motion sensitivity with E. Scaling was achieved by changing the distance between the subject and the screen, a method that also changes the sizes of the dots in the motion stimulus. Previous results along with the model presented in chapter 2 indicate that the rate of S_{min}^{100} scaling with E is very low (an E_2 of about 10), much lower than the scaling of visual acuity (an E_2 of about 2). Scaling by E_2 values larger than that for visual acuity can move dot sizes below visual acuity levels. Dot sizes below visual acuity exacerbate the Troxler effect, so measurements of temporal integration outside the fovea must include two types of scaling: one for the sizes of the structures in the image and one for the size of the motion stimulus.

7.5.2 Temporal response properties of motion detectors

The interactions of the temporal properties of the bi-local detector model components are also not easily measured. The actual presence of some sort of delay component may not be at all psychophysically testable. However, the use of an ISI in previous experiments has indicated that the front-end receptive field components will eventually stop firing if the spatial structure in the visual field is removed. The shape of the S_{max}^{100} curves in figure 6.1 for E=24 degrees also indicate that there is a distribution of transient and sustained detector types. If we assume that the temporal response of the detectors follows that of their front-end receptive fields, then the S_{max}^{100} curve at E=24 degrees in figure 6.1 would indicate that the transient type detectors serve the largest visible spatial displacements. At these largest displacements the front-end receptive field components stop firing after some time, even when the spatial structure remains present in the visual field. This can be deduced by considering that no ISI was used in any of the experiments presented here. Nevertheless, detectors of the largest S_s values at the tested non-foveal positions respond only when the frame rate is high enough, or conversely, when T_s is short enough.

A distribution of temporal response types is hinted at in the final figure presented in chapter 6. Transient (fast) type detectors should serve high

velocities while sustained (slow) type detectors server lower velocities and account for the T_S independence at smaller S_S values (e.g. see figure 2.4). The hypothesized distribution of temporal response types shown in figure 6.15 may be psychophysically testable. The critical-dimension results in chapter 5 and the simulation results in chapter 6 indicate that the receptive field size of the motion detector (and therefore the size of the front-end receptive field components) increases with preferred displacement size (S_S). Use of a spatially filtered version of the random dot stimulus (i.e. spatially filtering the random dot patterns) may therefore help isolate detector populations with specific sized receptive fields without incurring the following problem inherent in the use of displaced sinusoidal luminance gratings. When a sinusoidal luminance grating is spatially displaced, the movement direction indicated by the resulting spatio-temporal motion information reverses as the spatial displacement size reaches one half of the grating period. Motion detectors that respond to the displacement of the grating will therefore have an optimum displacement size less than that reversal point. The true optimum displacement size of the motion detector may be larger, but the grating stimulus paradigm contains the inherent limitation in terms of motion direction information.

7.5.3 Dependence of S_{max}^{100} and S_{min}^{100} on E

A more quantitative characterization of the variation of S_{max}^{100} and S_{min}^{100} would be valuable in refining the model because the linear magnification assumption is only approximate. Equation (6.10) may also indicate that a cosine fit would be more appropriate. More quantitative testing of S_{max}^{100} and S_{min}^{100} would provide the basis for a more precise evaluation of the relationship of the range of visible displacements with E. It may also be beneficial to develop more precise methods for fitting the model parameters to the visual performance of individual subjects, and to assess the variation of these parameters across the human population.

7.5.4 Critical stimulus dimensions

The measurement of critical dimensions as a function of S_S was entirely a qualitative process in chapter 5 and did not measure critical dimensions for spatial displacement sizes outside the fovea. A more quantitative measure of the dependence of the critical dimensions on S_S and E would be very valuable in further defining the distribution of motion detector receptive field sizes in the visual field. Such quantitative measures of the motion detector receptive field sizes could provide better guidance for self-organizational simulations as well as better interpretation of the results of those simulations. Moreover, the relationship between a motion detector's receptive field and preferred spatial displacement sizes can provide guidance in forming theoretical explanations of the formation of the detector array in terms of the available motion and object scales in visual space. The experimental paradigm presented in chapter 5 can be used unchanged to further the measurements of critical dimensions.

7.6 Other research directions

There are also a number of other research directions that would be valuable in extending our knowledge of motion perception. A few ideas along these lines are presented below.

7.6.1 Experimental Equipment

The special purpose device used to create the psychophysical stimuli used in the experiments presented here is one of only a few such devices and is in some ways limited. The device cannot vary pixel size, nor does it allow spatial filtering of the output images. Control of the spatial structure in the stimulus might allow a better separation of the spatial and temporal response characteristics of the motion detector population. The visual stimulus device was built using off-the-shelf electronic components that are by now somewhat outstripped by more modern technology. I would like to build a new version of the device that would fit on one or two printed circuit boards. The boards would be designed to fit in a personal computer and would be accompanied by a software control program similar to that used with the older machine. Such a design would allow a cheaper manufacturing process and a wider use of the stimulus paradigm, both in psychophysics and in neurophysiology. The construction of the next-generation stimulus device would be achieved through collaboration with the original designers in Utrecht.

7.6.2 3-D Motion

The stimulus paradigm developed for my dissertation work is directly extensible to three-dimensional motion perception by including time-varying stereo information. This avenue remained unexplored during my stay in the Netherlands due to lack of time. Such an extension of the stimulus might lead to a better understanding of the mechanisms of the perception of motion in three dimensions, and tests of whether the three-dimensional motion detection system is a direct calculation from the data in the visual field (independent from two dimensional motion) or a higher level combination of the outputs of the two-dimensional motion detectors.

7.6.3 Neurophysiology

One of my priorities is to initiate neurophysiological tests of the models proposed the previous chapters. The models of motion perception that I strive for are based on individual and populational characteristics of visual neural cells and visual neuroanatomy, but there have been no direct tests of neurophysiological responses to the stimulus employed in the experiments presented in my dissertation. No amount of psychophysics alone will allow us to precisely determine the biological mechanisms of motion perception or to make a direct connection between the psychophysical response to the stimulus and the response of visual cortical neurons to the stimulus. I would like to develop a collaboration with a neurophysiology laboratory interested in visual cortical recordings in primates, and especially in the biological computation of motion. Such a collaboration would provide the opportunity to directly test the response of cortical neurons in an animal visual system, preferably with the same or a similar battery of tests used in the psychophysical testing of human beings.

7.6.4 Computer vision theory

Computer vision scientists are interested in designing algorithms and systems that are effective in calculating visual information from environmental input. A useful approach to this endeavor is to ask ourselves what kinds of visual system properties are desirable and then to ponder the system design implications of the desired properties. The properties are often stated in the form of invariants of perception. One example is requiring detection of object forms without regard to spatial size, orientation, or position of the objects in the visual field. Such formulations have led to the creation of approaches to spatial image analysis called scale space methods. One result of that line of research is a set of theoretical spatial filters that meet the desired criteria. In fact, some cortical cells have been shown to have spatial receptive field shapes that match those predicted from the theoretically based deductions. Analysis of the same sort has led to a cataloging of invariant object shape properties that reveal themselves in the visual image of an object. The invariants are inherent to the shape of the object, but their perception depends on the position of the viewer and the relative motion between the object and viewer.

It is obvious that we can perceive object shapes when the objects are moving, and it is well known that a field of moving dots on a computer screen can be perceived as the surface of a three-dimensional object. Even when the dots in any single frame of the image sequence reveal no spatial structure, viewing the image sequence can result in a perception of the object that is absolutely unambiguous. Much of the research oriented toward image/shape invariance has been focused on the perception of form, but these invariance properties should also apply to the perception of objects in motion. In evolutionary terms, it is likely that motion perception systems appeared earlier than form perception systems, and Mother Nature's frugal engineering methods are (in)famous. This requires us to consider that the motion perception system participates in (or feeds into) the form perception system in some manner. (Note that artificial stabilization of an image on the retina results in a complete loss of perception of the image.) I would therefore like to study the theoretical implications of interactions of the motion detection system and the form detection system. For example, the motion detection system in primates begins early (area V1) in the processing stream. The input to V1 from the lateral geniculate body is almost exclusively center-surround, and although this is essentially one of the invariance preserving spatial filters, it is only one of them. Does the motion system use only these filters as input, or is the remainder of the scale-space filter set created before motion information is calculated? Does the motion system preserve the shape invariants resulting from the

shape of the object (that reveal themselves in the lighting conditions and viewpoint)? If only some of these invariants are preserved, which ones are preserved, and why? What form do these invariants hold after their transformation by the motion detecting system? Each of these questions is inherently interesting.

Moreover, the integrated model presented here provides some insight into the how the human visual system achieves such incredible efficiency and reliability. The spatial summation and temporal integration mechanisms provide powerful noise reduction (disambiguation) in the face of a number of sources of noise. Spatial summation actually conforms to an optimum predicted by signal processing theory. It would not be surprising if the temporal integration mechanism is also optimal in some sense. The structure of the motion perception system indicated by the integrated model presented here is therefore a good example for computer vision scientists to follow.

7.6.5 Self organization

A wide range of developmental experiments in animals indicate that the role of self organization in the formation of the visual system must be a major one. Some of the ideas presented in my dissertation are based on self-organization properties of visual systems, and part of my work to date has included explorations of these properties through simulation. I plan to continue exploring the role of developmental mechanisms in the formation of visual motion perception hardware.

7.7 Related perceptual mechanisms and phenomena

The 2+ years of research represented by this dissertation covers only a very small fraction of the phenomena that occur (or might occur) in the visual perception of motion. Here I briefly mention a number of those phenomena that are interesting to consider and are not addressed by the work presented here.

7.7.1 Priming

Priming refers to the idea that detectors of similar velocities can tell each other that something is about to pass. In the literature, the participating motion detectors are assumed to prefer the same direction and speed of motion and to be aligned on a path through the visual field. Each detector can then send warning information to the next detector on that path when it responds to the motion of an object through its receptive field. The warned detector is therefore more prepared to fire if and when the object passes through its receptive field, thereby increasing the ability of the motion detection system to detect motion in the presence of noise. In essence this a form of filtering of motion information.

7.7.2 Second order motion

Second order motion refers to visual motion that we can see but does not contain information that can be interpreted by the basic array of motion detectors. For example, if we set up a bar of motion noise on a mean luminance background and then move that bar across the visual field, many motion detectors will respond to the motion noise but there will be no coherent motion information extracted from it. However, we will very easily perceive the higher level motion indicated by the movement of the bar. Explanation of that percept requires some additional form of motion processing.

7.7.3 Cooperative/competitive interactions

There is a wide range of evidence that motion detectors can interact. Chapter 5 also presents possible evidence that the form perception system interacts with the motion system. The form of these interactions is important to our understanding of motion detection.

7.7.4 Attention

Attention is a very important modulator of perception. There is not a great deal known about the mechanism of the interaction, but the results are easy to see. The next time that you are traveling down the highway, look carefully at some small object in front of you. As you do you will no longer be able to attend to (see) the faster motion in your peripheral vision. Likewise, if you attend to that fast motion, you will no longer be able to judge the characteristics of the small object in front of you. Uncovering of that mechanism will be difficult because attention is a somewhat ill-defined phenomenon.

7.7.5 Long-range motion

The term long-range motion refers to a form of perceived motion that has different spatio-temporal characteristics than the motion perception that we can measure using random dot patterns. For example, if a large spot of luminance is positioned in the visual field, removed for a second and then placed in a new position, we can perceive a resulting motion of that spot. Under similar conditions in the random dot style experiments no motion would be perceived. The size of the spatial displacement and the length of the interstimulus interval (the time that the spot is not visible) that are conducive to the percept of motion are generally much larger than in the random dot pattern case, so the former is often called long-range motion while the latter is called short-range motion. The dichotomy implied by these terms is being challenged and defended in current research (some researchers feel that there is a continuous range of percepts rather than two separable ones).

7.8 Computer applications

The knowledge gained by mapping and modeling the visibility of motion in human vision should be useful in software/hardware system applications that need to display motion information. For example, the model may provide a basis for quantifying the display of information in terms of motion on a computer screen, or for predicting the visibility of apparent motion used in flight simulators and virtual environment facilities. Implementation of these ideas would require the use of direction-of-gaze tracking because of the foveo-centric nature of the distribution of motion detectors in the human visual system, but could provide an improved scene rendering process. Tracking eye positions would allow exploitation of the dependence on visual field eccentricity that occurs for visual acuity and for motion perception. The minimum perceivable grating frequency increases with angular distance from the point of gaze. This means that objects in the graphical scene away from the point of gaze can be rendered at a much coarser scale. The minimum and maximum perceivable pattern displacement size also increases with the angular distance from the point of gaze. Moreover, the receptive field size of motion detectors in human vision increases with the preferred spatial displacement size of the detector, so features that are moving fast relative to the point of gaze can also be rendered in lower detail. The combination of these visual system properties with eye tracking, while challenging at a systems design level, has the potential for a significant improvement in the utilization of computational power for graphics intensive applications.

7.9 Conclusions

This research provides some insight into how we can quantify the visibility of motion. The stimulus method developed for use in this research should provide a link between two-frame and multi-frame lifetime motion stimuli through the models presented in the previous chapters. The model components are not specific to apparent motion stimuli or to random dot stimuli, so more precise specification of the properties of the motion detector components may also allow application of the model to spatio-temporally continuous motion of other kinds. The integrated model for motion perception presented here provides a mechanistic basis for understanding how we perceive motion. The model can be tested both psychophysically and physiologically, and should be useful in building new computer vision systems as well as computer systems that must display motion information.

Bibliography

Adelson, E. H. and Bergen, J. R (1985) Spatiotemporal energy models for the perception of motion. *JOSA A*, **2**, 284-299.

Adler, B., Bock, O., and Grüsser, O.-J. (1981) Alpha-stripes: A spatial periodicity appearing in stroboscopically illuminated moving random dot patterns. *Vision Research*, **21**, 913-924.

Baker, C. L. Jr. and Braddick, O. J. (1985a) Eccentricity-dependent scaling of the limits for short-range apparent motion perception. *Vision Research*, **25**, 803-812.

Baker, C. L. Jr. and Braddick, O. J. (1985b) Temporal properties of the short-range process in apparent motion. *Perception*, **14**, 181-192.

Baker, C.L. Jr. and Braddick, O.J. (1985) Eccentricity-dependent scaling of the limits for short-range apparent motion perception. *Vision Research*, **25**, 803-812.

Ball, K. and Sekuler, R. (1979) Masking of motion by broadband and filtered directional noise. *Perception and Psychophysics*, **26**, 206-214.

Ball, K. and Sekuler, R. (1982) A specific and enduring improvement in visual motion discrimination. *Science*, **218**, 697-698.

Bischof, W. F. and Di Lollo, V. (1990) Perception of directional sampled motion in relation to displacement and spatial frequency: evidence of a unitary motion system. *Vision Research*, **30**, 1341-1362.

Bischof, W. F. and Groner, M. (1985) Beyond the displacement limit: an analysis of short-range processes in apparent motion. *Vision Research*, **25**, 839-847.

Blakemore, C., and Campbell, F.W. (1969). On the existence of neurones in the human visual sytem selectively sensitive to the orientation and size of retinal images, *Journal of Physiology*, **203**, pp. 237-260.

Boulton, J. C. (1987) Two mechanisms for the detection of slow motion. *JOSA A*, **4**, 1634-1642.

Boulton, J. C. and Baker, C. L. Jr. (1991) Motion detection is dependent on spatial frequency not size. *Vision Research*, **31**, 77-87.

Braddick, O. J. (1973) A short-range process in apparent motion. *Vision Research*, **14**, 519-527.

Browne, S. F. and McKee, S. P. (1989) Motion interference in speed discrimination. *JOSA A*, **6**, 1112-1121.

Burr, D. C. (1981) Temporal summation of moving images by the human visual system, *Proceedings of the Royal Society of London*, **B 211**, 321-339.

Burr, D.C., Ross, J. and Morrone, M.C. (1986) Smooth and sampled motion, *Vision Research*, **26**, 643-652.

Chang, J. J. and Julesz, B. (1983) Displacement limits for spatial frequency filtered random-dot cinematograms in apparent motion. *Vision Research*, **23**, 1379-2385.

Drasdo, N., (1991) Neural substrates and threshold gradients of peripheral vision. In Vision and Visual Dysfunction, Vol. 5, Macmillan Press, pp. 251-265.

Emerson, R.C., Bergen, J.R. and Adelsen, E.H. (1992) Directionally selective complex cells and the computation of motion energy in cat visual cortex. *Vision Research*, **32**, 203-218.

Fleet, D.J. and Jepson, A.D. (1985) Spatiotemporal inseparability in early vision: centre-surround models and velocity selectivity. In Computational Intelligence, Volume I, pp. 89-102.

Fredericksen, R. E., Verstraten, F.A.J. and Van de Grind, W.A. (submitted), An analysis of the mechanism underlying temporal integration of motion information.

Fredericksen, R. E., Verstraten, F.A.J. and Van de Grind, W.A., (in press) Spatio-temporal characteristics of human motion perception, to appear in *Vision Research*.

Fredericksen, R. E., Verstraten, F.A.J. and Van de Grind, W.A., (submitted) Temporal integration of random dot apparent motion in human central vision.

Fredericksen, R.E., Verstraten, F.A.J. and Van de Grind, W.A., (submitted). An analysis of the mechanism underlying temporal integration of motion information.

Grüsser, O.-J. (1991) Impairment of perception and recognition of face, facial expression and gestures in schizophrenic children and adolescents. In Eggers, C. (Ed.): Schizophrenia and youth, Springer, Heidelberg, 100-118.

Grüsser, O.-J., Landis, T. (1991) Visual agnosias and other disturbances of visual perception and cognition. In Vision and Visual Dysfunction, Vol. 12, MacMillan Press, London.

Kelly, D.H. and Burbeck, C.A. (1984) Critical problems in spatial vision, in CRC Critical Reviews in Biomedical Engineering, **10:2**, 125-177.

Koenderink, J. J. (1988) Scale time. Biological Cybernetics, 58, 159-162.

Lappin, J.S. and Bell, H.H. (1976) The detection of coherence in moving random-dot patterns. *Vision Research*, **16**, 161-168.

Luhmann, H.J., Martínez Millán, L., and Singer, W. (1986) Development of horizontal intrinsic connections in cat striate cortex. *Experimental Brain Research*, **63**, 443-448.

Marshall, J.A. (submitted) Adaptive perceptual pattern recognition by self-organizing neural networks: Context, uncertainty, multiplicity and scale.

Mastronarde, D.N. (1987) Two classes of single-input X-cells in cat lateral geniculate nucleus. I. Receptive-field properties and classification of cells. *Journal of Neurophysiology*, **57:2**, pp. 357-380.

McKee, S. P. and Welch, L. (1985) Sequential recruitment in the discrimination of velocity. *JOSA A*, **2**, 243-251.

Mikami, A., Newsome, W.T., and Wurtz, R. H., (1986) Motion selectivity in macaque visual cortex. I. Spatiotemporal range of directional interactions in MT and V1. *Journal of Neurophysiology*, **55:6**, pp. 1308-1327.

Mikami, A., Newsome, W.T., and Wurtz, R. H., (1986) Motion selectivity in macaque visual cortex. II. Mechanisms of direction and speed selectivity in extrastriate area MT. *Journal of Neurophysiology*, **55:6**, pp. 1328-1339.

Mikami, A., Newsome, W.T., and Wurtz, R. H., (1986) Motion selectivity in macaque visual cortex. III. Psychophysics and physiology of apparent motion. *Journal of Neurophysiology*, **55:6**, pp. 1340-1351.

Morgan, M. J. and Ward, R. (1980) Conditions for motion flow in dynamic visual noise. *Vision Research*, **20**, pp. 431-435.

Morgan, M.J. (1979) Perception of continuity in stroboscopic motion: A temporal frequency analysis, *Vision Research*, **19**, 491-500.

Nakayama, K. and Silverman, G. H. (1984) Temporal and spatial characteristics of the upper displacement limit for motion in random dots. *Vision Research*, **24**, 293-299.

Nakayama, K. and Tyler, C. W. (1981) Psychophysical isolation of movement sensitivity by removal of familiar position cues. *Vision Research*, **21**, 427-433.

Poggio, T. and Reichardt, W. (1973) Considerations on models of movement detection. *Kybernetik*, **13**, 223-227.

Proakis, J.G., Digital Communications, McGraw Hill Book Co., 1983.

Reichardt, W. (1961) Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. In Sensory Communication, Rosenblith, ed. (Wiley, New York, 1961), pp. 303-317.

Reid, R.C., Soodak, R.E. and Shapley, R.M. (1991) Directional selectivity and spatiotemporal structure of receptive fields of simple cells in cat striate cortex. *Journal of Neurophysiology*, **66:2**, pp. 505-529.

Robson, J.G., (1980) "Neural Images: The physiological basis of spatial vision," in Visual Coding and Adaptability, Ed. C.S. Harris, LEA, Hillsdale, New Jersey.

Santen, J. P. H. van and Sperling, G. (1985) Elaborated Reichardt detectors. *JOSA A*, **2**, 300-321.

Saul, A.B. and Humphrey, A.L. (1990) Spatial and temporal response properties of lagged and non-lagged cells in cat lateral geniculate nucleus. *Journal of Neurophysiology*, **64:1**, pp. 206-224.

Scobey, R. P. and Kan, P. L. E. (1991) A horizontal stripe of displacement sensitivity in the human visual field. *Vision Research*, **31**, 99-109.

Snowden, R. J. and Braddick, O. J. (1989a) Extension of displacement limits in multiple-exposure sequences of apparent motion. *Vision Research*, **29**, 1777-1787.

Snowden, R. J. and Braddick, O. J. (1989a) Extension of displacement limits in multiple-exposure sequences of apparent motion. *Vision Research*, **29**, 1777-1787.

Snowden, R. J. and Braddick, O. J. (1989b) The combination of motion signals over time. *Vision Research*, **29**, 1621-1630.

Snowden, R. J. and Braddick, O. J. (1990) Differences in the processing of short-range apparent motion at small and large displacements. *Vision Research*, **30**, 1211-1222.

Snowden, R. J. and Braddick, O. J. (1991) The temporal integration and resolution of velocity signals. *Vision Research*, **5**, 907-914.

Stremler, F.G. (1982) Introduction to communication systems, 2nd Ed., Addison-Wesley Publishing Company, Inc., Reading, Massachusetts.

Van Doorn, A.J., and Koenderink, J. J. (1982) Spatial properties of the visual detectability of moving spatial white noise. *Experimental Brain Research*, **45**, 189-198.

Van Doorn, A.J., and Koenderink, J. J. (1982) Temporal properties of the visual detectability of moving spatial white noise. *Experimental Brain Research*, **45**, 179-188.

Van Doorn, A.J., and Koenderink, J. J. (1984) Spatiotemporal integration in the detection of coherent motion. *Vision Research*, **24**, 47-53.

Van Doorn, A.J., and Koenderink, J. J. (1985) Perception of movement and correlation in stroboscopically presented noise patterns. *Perception*, **14**, 209-224.

Van de Grind, W. A., and Grüsser, O.-J. (1973) Temporal transfer properties of the afferent visual system, in Handbook of Sensory Physiology, Volume VII/3A, R. Jung, editor.

Van de Grind, W. A., Van Doorn, A. J., and Koenderink, J. J. (1983) Detection of coherent movement in peripherally viewed random dot patterns. *JOSA*, **73**, 1674-1683.

Van de Grind, W. A., Koenderink, J. J. and Van Doorn, A. J. (1986) The distribution of human motion detector properties in th monocular visual field. *Vision Research*, **26**, 797-810.

Van de Grind, W. A., Koenderink, J. J., and Van Doorn, A. J. (1987) Influence of contrast on foveal and peripheral detection of coherent motion in moving random-dot patterns. *JOSA A*, **4**, 1643-1652.

Van de Grind, W. A., Koenderink, J. J., Van Doorn, A.J., Milders, M.V., and Voerman, H. (in press) Inhomogeneity and anisotropies for motion detection in the monocular visual field of human observers. To appear in *Vision Research*.

Wattam-Bell, J. (1991) Development of motion-specific cortical responses in infancy. *Vision Research*, **31**, 287-297.

Wattam-Bell, J. (1992) Development of maximum displacement limits for discrimination of motion direction in infancy. *Vision Research*, **32**, 621-630.

Williams, D. W. and Sekuler, R. (1984) Coherent global motion percepts from stochastic local motions. *Vision Research*, 24, 55-62.

Williams, W., Phillips, G. and Sekuler, R. (1986) Hysteresis in the perception of motion direction as evidence for neural cooperativity. *Nature*, **324**, 253-255.

Wilson, H. R., Levi, D., Maffei, L., Rovamo, J. and DeValois, R. (1990) The perception of form: Retina to striate cortex. In Visual Perception: The Neurophysiological Foundations, Spillman and Werner, eds. (Academic Press, Inc., San Diego, 1990) pp. 231-272.